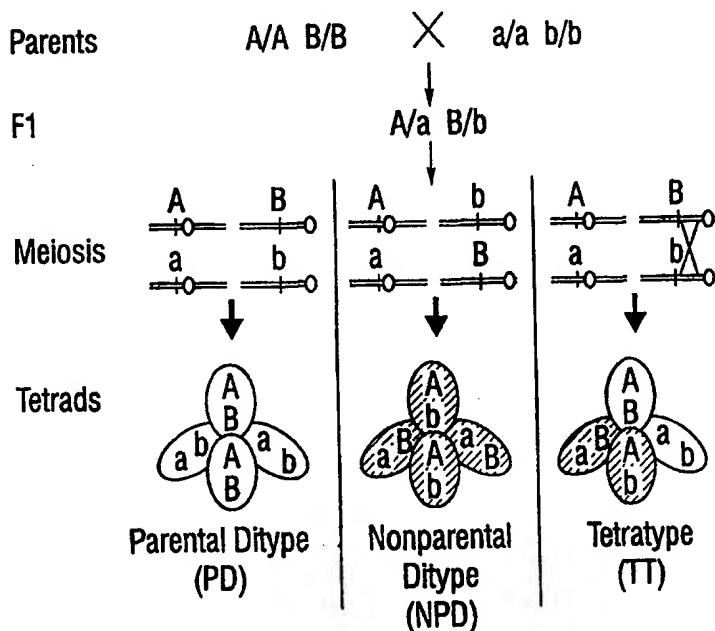




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|  |  |  |   |
|--|--|--|---|
| (51) International Patent Classification 6 :<br><b>C12N 15/82, 1/21, 5/10, A01H 5/00</b>   |  | A1   | (11) International Publication Number: <b>WO 98/55637</b><br>(43) International Publication Date: 10 December 1998 (10.12.98) |
| (21) International Application Number: <b>PCT/US98/11288</b>   |  | (74) Agent: HANSON, Robert, E.; Arnold, White & Durkee, P.O. Box 4433, Houston, TX 77210 (US).   |   |
| (22) International Filing Date: 3 June 1998 (03.06.98)   |  | (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). |   |
| (30) Priority Data:<br>60/048,451 3 June 1997 (03.06.97) US<br>60/073,741 5 February 1998 (05.02.98) US                                    |  | (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications<br>US 60/048,451 (CON) 3 June 1997 (03.06.97)<br>US 60/073,741 (CON) 5 February 1998 (05.02.98)   |   |
| (71) Applicant (for all designated States except US): ARCH DEVELOPMENT CORPORATION [US/US]; 1011 East 58th Street, Chicago, IL 60637 (US). |  | (72) Inventors; and<br>(75) Inventors/Applicants (for US only): COPENHAVER, Gregory [US/US]; 1028 S. Elmwood, Oakpark, IL 60304 (US). PREUSS, Daphne [US/US]; 5807 S. Dorchester No. 2E, Chicago, IL 60637 (US).   |   |

(54) Title: PLANT ARTIFICIAL CHROMOSOME (PLAC) COMPOSITIONS AND METHODS



## (57) Abstract

The present invention provides for the identification and cloning of functional plant centromeres in *Arabidopsis*. This will permit construction of stably inherited plant artificial chromosomes (PLACs) which can serve as vectors for the construction of transgenic plant and animal cells. In addition, information on the structure and function of these regions will prove valuable in isolating additional centromeric and centromere related genetic elements and polypeptides from other species.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |                                       |    |   |    |                          |
|----|--------------------------|----|---------------------------------------|----|---|----|--------------------------|
| AL | Albania                  | ES | Spain                                 | LS | Lesotho                                   | SI | Slovenia                 |
| AM | Armenia                  | FI | Finland                               | LT | Lithuania                                 | SK | Slovakia                 |
| AT | Austria                  | FR | France                                | LU | Luxembourg                                | SN | Senegal                  |
| AU | Australia                | GA | Gabon                                 | LV | Latvia                                    | SZ | Swaziland                |
| AZ | Azerbaijan               | GB | United Kingdom                        | MC | Monaco                                    | TD | Chad                     |
| BA | Bosnia and Herzegovina   | GE | Georgia                               | MD | Republic of Moldova                       | TG | Togo                     |
| BB | Barbados                 | GH | Ghana                                 | MG | Madagascar                                | TJ | Tajikistan               |
| BE | Belgium                  | GN | Guinea                                | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan             |
| BF | Burkina Faso             | GR | Greece                                | ML | Mali                                      | TR | Turkey                   |
| BG | Bulgaria                 | HU | Hungary                               | MN | Mongolia                                  | TT | Trinidad and Tobago      |
| BJ | Benin                    | IE | Ireland                               | MR | Mauritania                                | UA | Ukraine                  |
| BR | Brazil                   | IL | Israel                                | MW | Malawi                                    | UG | Uganda                   |
| BY | Belarus                  | IS | Iceland                               | MX | Mexico                                    | US | United States of America |
| CA | Canada                   | IT | Italy                                 | NE | Niger                                     | UZ | Uzbekistan               |
| CF | Central African Republic | JP | Japan                                 | NL | Netherlands                               | VN | Viet Nam                 |
| CG | Congo                    | KE | Kenya                                 | NO | Norway                                    | YU | Yugoslavia               |
| CH | Switzerland              | KG | Kyrgyzstan                            | NZ | New Zealand                               | ZW | Zimbabwe                 |
| CI | Côte d'Ivoire            | KP | Democratic People's Republic of Korea | PL | Poland                                    |    |                          |
| CM | Cameroon                 | KR | Republic of Korea                     | PT | Portugal                                  |    |                          |
| CN | China                    | KZ | Kazakhstan                            | RO | Romania                                   |    |                          |
| CU | Cuba                     | LC | Saint Lucia                           | RU | Russian Federation                        |    |                          |
| CZ | Czech Republic           | LI | Liechtenstein                         | SD | Sudan                                     |    |                          |
| DE | Germany                  | LK | Sri Lanka                             | SE | Sweden                                    |    |                          |
| DK | Denmark                  | LR | Liberia                               | SG | Singapore                                 |    |                          |
| EE | Estonia                  |    |                                       |    |   |    |                          |

DESCRIPTION

## PLANT ARTIFICIAL CHROMOSOME (PLAC) COMPOSITIONS AND METHODS

The government owns rights in the present invention pursuant to U.S. Department of Agriculture Grant No. 96-35304-3491 and Grant No. DE-FC05-920R22072 from the Consortium for Plant Biotechnology. This application claims the priority of U.S. Provisional Patent Application Serial No. 60/048,451, filed June 3, 1997; and U.S. Provisional Patent Application Serial No. 60/073,741, filed February 5, 1998, both of the disclosures of which are specifically incorporated herein by reference in their entirety.

**I. Field of the Invention**

The present invention relates generally to the field of molecular biology. More particularly, it concerns the construction and use of plant artificial chromosomes.

**10 II. Description of Related Art**

Two general approaches are used for introduction of new genetic information ("transformation") into cells. One approach is to introduce the new genetic information as part of another DNA molecule, referred to as a "vector," which can be maintained as an independent unit (an episome) apart from the chromosomal DNA molecule(s). Episomal vectors contain all the necessary DNA sequence elements required for DNA replication and maintenance of the vector within the cell. Many episomal vectors are available for use in bacterial cells (for example, see Maniatis *et al.*, 1982). However, only a few episomal vectors that function in higher eukaryotic cells have been developed. The available higher eukaryotic episomal vectors are based on naturally occurring viruses and most function only in mammalian cells (Willard, 1997). In higher plant systems the only known double-stranded DNA viruses that replicate through a double-stranded intermediate upon which an episomal vector could be based is the gemini virus, although the gemini virus is limited to an approximately 800 bp insert. Although an episomal plant vector based on the Cauliflower Mosaic Virus has been developed, its capacity to carry new genetic information also is limited (Brisson *et al.*, 1984).

25 The other general method of genetic transformation involves integration of introduced DNA sequences into the recipient cell's chromosomes, permitting the new information to be

replicated and partitioned to the cell's progeny as a part of the natural chromosomes. The most common form of integrative transformation is called "transfection" and is frequently used in mammalian cell culture systems. Transfection involves introduction of relatively large quantities of deproteinized DNA into cells. The introduced DNA usually is broken and joined together in various combinations before it is integrated at random sites into the cell's chromosome (see, for example Wigler *et al.*, 1977). Common problems with this procedure are the rearrangement of introduced DNA sequences and unpredictable levels of expression due to the location of the transgene in the genome or so called "position effect variation" (Shingo *et al.*, 1986). Further, unlike episomal DNA, integrated DNA cannot normally be precisely removed. A more refined form of integrative transformation can be achieved by exploiting naturally occurring viruses that integrate into the host's chromosomes as part of their life cycle, such as retroviruses (see Cepko *et al.*, 1984). In mouse, homologous integration has recently become common, although it is significantly more difficult to use in plants (Lam *et al.* 1996).

The most common genetic transformation method used in higher plants is based on the transfer of bacterial DNA into plant chromosomes that occurs during infection by the phytopathogenic soil bacterium *Agrobacterium* (see Nester *et al.*, 1984). By substituting genes of interest for the naturally transferred bacterial sequences (called T-DNA), investigators have been able to introduce new DNA into plant cells. However, even this more "refined" integrative transformation system is limited in three major ways. First, DNA sequences introduced into plant cells using the *Agrobacterium* T-DNA system are frequently rearranged (see Jones *et al.*, 1987). Second, the expression of the introduced DNA sequences varies between individual transformants (see Jones *et al.*, 1985). This variability is presumably caused by rearranged sequences and the influence of surrounding sequences in the plant chromosome (*i.e.*, position effects), as well as methylation of the transgene. A third drawback of the *Agrobacterium* T-DNA system is the reliance on a "gene addition" mechanism: the new genetic information is added to the genome (*i.e.*, all the genetic information a cell possesses) but does not replace information already present in the genome.

One attractive alternative to commonly used methods of transformation is the use of an artificial chromosome. Artificial chromosomes are man-made linear or circular DNA molecules constructed from essential cis-acting DNA sequence elements that are responsible for the proper replication and partitioning of natural chromosomes (see Murray *et al.*, 1983).

The essential elements are: (1) Autonomous Replication Sequences (ARS) (these have properties of replication origins, which are the sites for initiation of DNA replication), (2) Centromeres (site of kinetochore assembly and responsible for proper distribution of replicated chromosomes at mitosis or meiosis), and (3) Telomeres (specialized structures at the ends of 5 linear chromosomes that function to stabilize the ends and facilitate the complete replication of the extreme termini of the DNA molecule).

At present, the essential chromosomal elements for construction of artificial chromosomes have been precisely characterized only from lower eukaryotic species. ARSs have been isolated from unicellular fungi, including *Saccharomyces cerevisiae* (brewer's yeast) 10 and *Schizosaccharomyces pombe* (see Stinchcomb *et al.*, 1979 and Hsiao *et al.*, 1979). ARSs behave like replication origins allowing DNA molecules that contain the ARS to be replicated as an episome after introduction into the cell nuclei of these fungi. Plasmids containing these sequences replicate, but in the absence of a centromere they are partitioned randomly into daughter cells.

15 Artificial chromosomes have been constructed in yeast using the three cloned essential chromosomal elements. Murray *et al.*, 1983, disclose a cloning system based on the *in vitro* construction of linear DNA molecules that can be transformed into yeast, where they are maintained as artificial chromosomes. These yeast artificial chromosomes (YACs) contain cloned genes, origins of replication, centromeres and telomeres and are segregated in daughter 20 cells with high affinity when the YAC is at least 100 kB in length. Smaller CEN containing vectors may be stably segregated, however, when in circular form.

None of the essential components identified in unicellular organisms, however, function 25 in higher eukaryotic systems. For example, a yeast CEN sequence will not confer stable inheritance upon vectors transformed into higher eukaryotes. While such DNA fragments can be readily be introduced, they do not stably exist as episomes in the host cell. This has seriously hampered efforts to produce artificial chromosomes in higher organisms.

In one case, a plant artificial chromosome was discussed (Richards *et al.*, U.S. Patent 30 No. 5,270,201). However, this vector was based on plant telomeres, as a functional plant centromere was not disclosed. While telomeres are important in maintaining the stability of chromosomal termini, they do not encode the information needed to ensure stable inheritance of

an artificial chromosome. It is well documented that centromere function is crucial for stable chromosomal inheritance in almost all eukaryotic organisms (reviewed in Nicklas 1988). For example, broken chromosomes that lack a centromere (acentric chromosomes) are rapidly lost from cell lines, while fragments that have a centromere are faithfully segregated. The 5 centromere accomplishes this by attaching, via centromere binding proteins, to the spindle fibers during mitosis and meiosis, thus ensuring proper gene segregation during cell divisions.

In contrast to the detailed studies done in *S. cerevisiae* and *S. pombe*, little is known about the molecular structure of functional centromeric DNA of higher eukaryotes. Ultrastructural studies indicate that higher eukaryotic kinetochores, which are specialized 10 complexes of proteins that form on the chromosome during late prophase, are large structures (mammalian kinetochore plates are approximately 0.3  $\mu\text{m}$  in diameter) which possess multiple microtubule attachment sites (reviewed in Rieder, 1982). It is therefore possible that the centromeric DNA regions of these organisms will be corresponding large, although the minimal amount of DNA necessary for centromere function may be much smaller.

15 While the above studies have been useful in elucidating the structure and function of centromeres, they have failed to provide a cloned, functional centromere from a higher eukaryotic organism. The extensive literature indicating both the necessity of centromeres for stable inheritance of chromosomes, and the non-functionality of yeast centromeres in higher organisms, demonstrate that cloning of a functional centromere from a higher eukaryote is a necessary first step in the production of artificial chromosomes suitable for use in higher plants 20 and animals. The production of artificial chromosomes with centromeres which function in higher eukaryotes would overcome many of the problems associated with the prior art and represent a significant breakthrough in biotechnology research.

## SUMMARY OF THE INVENTION

25 The current invention overcomes deficiencies in the prior art by providing methods for obtaining a functional plant centromere and uses therefor. More particularly, the present invention provides for the production of a stably inherited plant artificial chromosome.

In a first embodiment, there is provided a method for the identification plant 30 centromeres. Briefly, tetrad analysis measures the recombination frequency between genetic makers and a centromere by analyzing all four products of individual meiosis. A particular

advantage arises from the quartet (qrt 1) mutation in *Arabidopsis*, which causes the four products of pollen mother cell meiosis in *Arabidopsis* to remain attached. When used to pollinate a flower, one tetrad can result in the formation of four seeds, and the plants from these seeds can be analyzed genetically. With unordered tetrads, however, such as those produced by 5 *Arabidopsis*, genetic mapping using tetrad analysis requires that two markers be scored simultaneously.

In another embodiment, the present invention provides new plant artificial chromosomes (PLACs) and DNA fragments for the creation thereof. In a preferred embodiment the PLAC will have functional sequences which include telomeres, a plant and / or 10 other autonomous replicating sequence, a centromere, and selectable markers which confer a growth advantage under particular conditions to plant cells carrying the marker, thereby allowing identification of plants, plants cells or cells from any other organism of interest containing the PLAC. A selectable marker may in particular embodiments of the invention function in bacterial cells. The PLAC also may contain "negative" selectable markers which 15 confer susceptibility to an antibiotic, herbicide or other agent, thereby allowing for selection against plants, plant cells or cells of any other organism of interest containing a PLAC. The PLAC also may include genes which control the copy number of the PLAC within a cell. One or more structural genes also may be included in the PLAC. Specifically contemplated as being useful will be as many structural genes as may be inserted into the PLAC while still 20 maintaining a functional vector. This may include one, two, three, four, five, six, seven, eight, nine or more structural genes.

In another embodiment, the invention provides methods for expressing foreign genes in plants, plant cells or cells of any other organism of interest. The foreign genes may be from any organism, including plants, animals and bacteria. It is further contemplated that PLACs 25 could be used to simultaneously transfer multiple foreign genes to a plant comprising entire biochemical or regulatory pathways. In yet another embodiment of the invention, it is contemplated that the PLACs can be used as DNA cloning vectors. Such a vector could be used in plant and animal sequencing projects. The current invention may be of particular use in the cloning of sequences which are "unclonable" in yeast and bacteria, but which may be easier 30 to clone in a plant based system.

In still yet another embodiment of the invention, it is contemplated that the PLACs disclosed herein may be used clone functional segments of DNA such as origins of DNA replication, telomeres, telomere associated genes, nuclear matrix attachment regions (MARs), scaffold attachment regions (SARs), boundary elements, enhancers, silencers, promoters, 5 recombinational hot-spots and centromeres. This embodiment may be carried out by cloning DNA into a defective PLAC which is deficient for one or more type of functional elements. Sequences which complemented such deficient elements would cause the PLAC to be stably inherited. A selectable marker on the PLAC could then be used to select for viable PLAC containing cells which contain cloned functional elements of the type that were non-functional 10 in the defective PLAC.

In still yet another embodiment of the invention, the sequences disclosed herein may be used for the isolation of centromeric sequences from other plant species including agriculturally important species such as *Brassica* species (Broccoli, Cauliflower, Mustard, *etc.*), and monocots such as wheat and corn. Methods for isolating centromeric sequences using the 15 sequences disclosed herein are well known in the art and are based on shared homology of the centromeric sequences.

In still yet another embodiment of the invention, the artificial chromosome vectors described herein may be used to perform efficient gene replacement studies. At present, gene replacement has been detected on only a few occasions in plant systems and has only been 20 detected at low frequency in mammalian tissue culture systems (see Thomas *et al.*, 1986; Smithies *et al.*, 1985). The reason for this is the high frequency of illegitimate nonhomologous recombination events relative to the frequency of homologous recombination events (the latter are responsible for gene replacement). Artificial chromosomes may participate in homologous recombination preferentially. Since the artificial chromosomes remain intact upon delivery, no 25 recombinogenic broken ends will be generated to serve as substrates for the extremely efficient illegitimate recombination machinery. Thus, the artificial chromosome vectors disclosed by the present invention will be stably maintained in the nucleus through meiosis and available to participate in homology-dependent meiotic recombination. In addition, because in principle, artificial chromosomes of any length could be constructed using the teaching of the present 30 invention, the vectors could be used to introduce extremely long stretches of DNA from the same or any other organism into cells. Specifically contemplated inserts include those from

about several base pairs to one hundred megabase pairs, including about 1 kb, 25 kB, 50 kB, 100 kB, 125 kB, 150 kB, 200 kB, 300 kB, 400 kB, 500 kB, 600 kB, 700 kB, 800 kB, 900 kB, 1 MB, 1.25 Mb, 1.5 Mb, 2 Mb, 3 Mb, 5 Mb, 10 Mb, 25 Mb, 50 Mb and 100 Mb.

In still yet another embodiment, the present invention relates to the construction of 5 artificial chromosome vectors for the genetic transformation of plant cells, processes for their preparation, uses of the vectors, and organisms transformed by them. Standard reference works setting forth the general principles of recombinant DNA technology include Lewin, 1985. Other works describe methods and products of genetic engineering; see, e.g., Maniatis *et al.*, 1982; Watson *et al.*, 1983; Setlow *et al.*, 1979; and Dillon *et al.*, 1985.

10

#### BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

15

FIG. 1: Centromere mapping with unordered tetrads: A cross of two parents (AABB x aabb), in which "A" is on the centromere of one chromosome, and "B" is linked to the centromere of a second chromosome. At meiosis, the A and B chromosomes assort independently, resulting in equivalent numbers of parental ditype (PD) and nonparental ditype (NPD) tetrads (recombinant progeny are shown in gray). Tetratype tetrads (TT) result only 20 from a crossover between "B" and the centromere.

25

FIG. 2: Low resolution map location of *Arabidopsis* centromeres. Trisomic mapping was used to determine the map position of centromeres on four of the five *Arabidopsis* chromosomes (Koornneef 1983; Sears *et al.*, 1970). For chromosome 4, useful trisomic strains were not obtained. With the methods of Koornneef and Sears *et al.*, 1983. (which rely on low-resolution deletion mapping) the centromere on chromosome 1 was found to lie between the two visible markers, *tt1* and *ch1*, that are separated by 5 cM. Centromere positions on the other chromosomes are mapped to a lower resolution.

FIG. 3A-E: Map location of *Arabidopsis thaliana* centromeres on chromosomes 1-5. Centromeres were mapped by tetrad analysis as described below. Genetic markers used to map

the centromeres are displayed above the chromosomes and are designated by their cM values (all cM values were derived from the Lister and Dean recombinant inbred map at the web site: [http://genome-www3.stanford.edu/atdb\\_welcome.html](http://genome-www3.stanford.edu/atdb_welcome.html)).

FIG. 4: Seed stock used for tetrad analysis in *Arabidopsis thaliana*. The individual 5 strains are identified by the strain number. The tetrad member number indicates the tetrad source (i.e. T1 indicates seeds from tetrad number 1, and the numbers -1, -2, -3, or -4 indicate individual members of the tetrad). The strains listed have been deposited with the *Arabidopsis* 10 Biological Resources Center (ABRC) at Ohio State University under the name of Daphne Preuss, or where seed set has not yet occurred, will be deposited when collected (indicated in 15 column 3).

FIG. 5: Marker information for centromere mapping. DNA polymorphisms used to 15 localize the centromeres are indicated by chromosome (Column 1). The name of each marker is shown in Column 2, and the marker type in Column 3. CAPS (Co-dominant Amplified Polymorphic Sites) are markers that can be amplified with PCR and detected by digesting with 20 the appropriate restriction enzyme (also indicated in Column 3). SSLPs (Simple Sequence Length Polymorphisms) detect polymorphisms by amplifying different length PCR products. Column 4 notes if the marker is available on public web sites (<http://genome-www.stanford.edu/Arabidopsis>). For those markers that are not available on public web sites the sequences of the forward and reverse primers used to amplify the marker are listed in 25 columns 5 and 6, respectively, and given in SEQ ID NO:1 through SEQ ID NO:44.

FIG. 6: Scoring PCR-based markers for tetrad analysis. The genotype of the progeny from one pollen tetrad (T2) was determined for two genetic markers (SO392 and nga76). Analysis of the four progeny plants (T2-1 through T2-4) using PCR and gel electrophoresis allows the genotype of the plant to be determined, and the genotype of the pollen parent to be 25 inferred.

FIGs. 7A-7H: Exemplary PLAC vectors: The vectors shown in FIG. 7A, FIG. 7B, FIG. 7E, and FIG. 7F have an *E. coli* origin of replication which can be high copy number, low copy number or single copy. In FIGs. 7A-7H, the vectors include a multiple cloning site which can contain recognition sequences for conventional restriction endonucleases with 4-8 bp 30 specificity as well as recognition sequences for very rare cutting enzymes such as, for example,

I-Ppo I, I-Ceu I, PI-Tli I, PI-Psp I, Not I, and PI Sce I. In FIG. 7A, FIG. 7B, FIG. 7C, FIG. 7D, FIG. 7E, FIG. 7G, and FIG. 7H the centromere is flanked by Lox sites which can act as targets for the site specific recombinase Cre. FIG. 7A: An *E. Coli* plant circular shuttle vector with a plant ARS. FIG. 7B: A plant circular shuttle vector without a plant ARS. The vector relies on 5 a plant origin of replication function found in other plant DNA sequences such as selectable markers. FIG. 7C: A yeast-plant circular shuttle vector with a plant ARS. The yeast ARS is included twice, once on either side of the multiple cloning site to ensure that large inserts are stable. FIG. 7D: A yeast-plant circular shuttle vector without a plant ARS. The vector relies on a plant origin of replication function found in other plant DNA sequences such as selectable 10 markers. The yeast ARS is included twice, once on either side of the multiple cloning site to ensure that large inserts are stable. FIG. 7E: An *E. Coli* -*Agrobacterium* - Plant circular shuttle vector with a plant ARS. Vir functions for T-DNA transfer would be provided in trans by using the appropriate *Agrobacterium* strain. FIG. 7F: An *E. Coli* - *Agrobacterium* - plant circular shuttle vector without a plant ARS. The vector relies on the plant origin of replication function 15 found in other plant DNA sequences such as selectable markers. Vir functions for T-DNA transfer would be provided in trans by using the appropriate *Agrobacterium* strain. FIG. 7G: A linear plant vector with a plant ARS. The linear vector could be assembled in vitro and then transferred into the plant by, for example, mechanical means such as microprojectile bombardment, electroporation, or PEG-mediated transformation. FIG. 7H: A linear plant 20 vector without a plant ARS. The linear vector could be assembled in vitro and then transferred into the plant by, for example, mechanical means such microprojectile bombardment, electroporation, and PEG-mediated transformation.

#### DETAILED DESCRIPTION OF THE INVENTION

The prior art has failed to provide a centromere which is functional in plants. This 25 failure is exemplified by the general lack of detailed information in the art regarding the centromeres of multicellular organisms in general. To date, the most extensive and reliable characterization of centromere sequences has come from studies of lower eukaryotes such as *S. cerevisiae* and *S. pombe*, where the ability to analyze centromere functions has provided a clear picture of the essential DNA sequences. The *S. cerevisiae* centromere consists of three 30 essential regions, CDEI, CDEII, and CDEIII, totaling only 125 bp, or approximately 0.006 to 0.06% of each yeast chromosome (Carbon *et al.*, 1990; Bloom 1993). *S. pombe* centromeres

are between 40 and 100 kB in length and consist of repetitive elements that comprise 1 to 3% of each chromosome (Baum *et al.*, 1994). Subsequent studies, using tetrad analysis to follow the segregation of artificial chromosomes, demonstrated that less than 1/5 of the naturally occurring *S. pombe* centromere is sufficient for centromere function (Baum *et al.*, 1994).

5        In contrast, the centromeres of mammals and other higher eukaryotes are poorly defined. Although DNA fragments that hybridize to centromeric regions in higher eukaryotes have been identified, little is known regarding the functionality of these sequences (see Tyler-Smith *et al.*, 1993). In many cases centromere repeats correlate with centromere location, with probes to the repeats mapping both cytologically and genetically to centromere regions. Many  
10      of these sequences are tandemly-repeated satellite elements and dispersed repeated sequences in arrays ranging from 300 kB to 5000 kB in length (Willard 1990). To date, only one of these repeats, a 171 bp element known as the alphoid satellite, has been shown by *in situ* hybridization to be present at each human centromere (Tyler-Smith *et al.*, 1993). Whether repeats themselves represent functional centromeres remains controversial, as other genomic  
15      DNA is required to confer inheritance upon a region of DNA (Willard, 1997). Alternatively, the positions of some higher eukaryotic centromeres have been estimated by analyzing the segregation of chromosome fragments. This approach is imprecise, however, because a limited set of fragments can be obtained, and because normal centromere function is influenced by surrounding chromosomal sequences (for example, see Koornneef, 1983; FIG. 2).

20        A more precise method for mapping centromeres that can be used in intact chromosomes is tetrad analysis (Mortimer *et al.*, 1981), which provides a functional definition of a centromere in its native chromosomal context. At present, the only centromeres that have been mapped in this manner are from unicellular eukaryotes, including the yeasts *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Kluyveromyces lactis* (Carbon *et al.*, 1990; Hegemann *et al.*, 1993). In these systems, accurate mapping of the centromeres made it possible to clone centromeric DNA, using a chromosome walking strategy (Clarke *et al.*, 1980). Subsequently, artificial chromosome assays were used to define more precisely the centromere sequences (Hegemann *et al.*, 1993; Baum *et al.*, 1994).

25        Attempts to develop a reliable centromeric assay in mammals have yielded ambiguous results. For example, Hadlaczky *et al.*, (1991) identified a 14 kB human fragment that can, at low frequency, result in *de novo* centromere formation in a mouse cell line. *In situ*

hybridization studies, however, have shown that this fragment is absent from naturally occurring centromeres, calling into question the reliability of this approach for testing centromere function (Tyler-Smith *et al.*, 1993). Similarly, transfection of alphoid satellites into cell lines results in the formation of new chromosomes, yet these chromosomes also contain 5 host sequences that could contribute centromere activity (Haaf *et al.*, 1992; Willard, 1997). Further, the novel chromosomes can have alphoid DNA spread throughout their length yet have only a single centromeric constriction, indicating that a block of alphoid DNA alone may be insufficient for centromere function (Tyler-Smith *et al.*, 1993).

Although plant centromeres can be visualized easily in condensed chromosomes, they 10 have not been characterized as extensively as centromeres from yeast or mammals. Genetic characterization has relied on segregation analysis of chromosome fragments, and in particular on analysis of trisomic strains that carry a genetically marked, telocentric fragment (for example, see Koornneef 1983; FIG. 2). In addition, repetitive elements have been identified that are either genetically (Richards *et al.*, 1991) or physically (Alfenito *et al.*, 1993; 15 Maluszynska *et al.*, 1991) linked to a centromere. In no case, however, has the functional significance of these sequences been tested.

Cytology in *Arabidopsis thaliana* has served to correlate centromere structure with 20 repeat sequences. A fluorescent dye, DAPI, allows visualization of centromeric chromatin domains in metaphase chromosomes. A fluorescence *in situ* hybridization (FISH) probe based on 180 bp pAL1 repeat sequences colocalized with the DAPI signature near the centromeres of all five *Arabidopsis* chromosomes (Maluszynska *et al.*, 1991; Martinez-Zapater *et al.*, 1986), however this repeat probe also hybridizes within noncentromeric regions of the chromosomes. 25 Although a functional role for pAL1 has been proposed, more recent studies have failed to detect this sequence near the centromeres in species closely related to *Arabidopsis thaliana* (Maluszynska *et al.*, 1993). These results are particularly troubling because one of the species tested, *A. pumila*, is thought to be an amphidiploid, derived from a cross between *A. thaliana* and another close relative (Maluszynska *et al.*, 1991; Price *et al.*, 1995). Another repetitive 30 sequence, pAtT12, has been genetically mapped to within 5 cM of the centromere on chromosome 1 and to the central region of chromosome 5 (Richards *et al.*, 1991), although its presence on other chromosomes has not been established. Like pAL1, a role for pAtT12 in centromere function remains to be demonstrated.

Due to the fact that kinetochores constitute a necessary link between centromeric DNA and the spindle apparatus, the proteins that are associated with these structures recently have been the focus of intense investigation (Bloom 1993; Earnshaw 1991). Human autoantibodies that bind specifically in the vicinity of the centromere have facilitated the cloning of 5 centromere-associated proteins (CENPs, Rattner 1991), and at least one of these proteins belongs to the kinesin superfamily of microtubule-based motors (Yen 1991). Yeast centromere-binding proteins also have been identified, both through genetic and biochemical studies (Bloom 1993; Lechner *et al.*, 1991). Although the basic features of mitosis and meiosis (including spindle attachment, chromosome pairing, and chromosome separation) are conserved 10 among all eukaryotes, a comparison of the known centromere-binding proteins from yeast and humans reveals little similarity (Bloom, 1993; Earnshaw and Cooke, 1989).

The centromeres of *Arabidopsis thaliana* have been mapped using trisomic strains, where the segregation of chromosome fragments (Koornneef 1983) or whole chromosomes (Sears *et al.*, 1970) was used to localize four of the centromeres to within 5, 12, 17 and 38 cM, 15 respectively (FIG. 2). These positions have not been refined by more recent studies because the method is limited, not only by the difficulty in obtaining viable trisomic strains, but also by the inaccuracy of comparing recombination frequencies in chromosome fragments to those in intact chromosomes (Koornneef 1983). These factors introduce significant error into the calculated position of the centromere, and in *Arabidopsis*, where 1 cM corresponds roughly to 200 kB 20 (Koornneef 1987; Hwang *et al.*, 1991), this method did not map any of the centromeres with sufficient precision to make chromosome walking strategies practical.

### I. Tetrad Analysis

With tetrad analysis, the recombination frequency between genetic markers and a centromere can be measured directly (FIG. 1). This method requires analysis of all four 25 products of individual meiosis, and it has not been applied previously to multicellular eukaryotes because their meiotic products typically are dissociated. Identification of the *quartet* mutation makes tetrad analysis possible for the first time in a multicellular genetic model system (Preuss *et al.*, 1994). The *quartet* (qrt 1) mutation causes the four products of pollen mother cell meiosis in *Arabidopsis* to remain attached. When used to pollinate a flower, 30 one tetrad can result in the formation of four seeds, and the plants from these seeds can be analyzed genetically.

With unordered tetrads, such as those produced by *S. cerevisiae* or *Arabidopsis*, genetic mapping using tetrad analysis requires that two markers be scored simultaneously (Whitehouse 1950). Tetrads fall into different classes depending on whether the markers are in a parental (nonrecombinant) or nonparental (recombinant) configuration (FIG. 1). A tetrad with only 5 nonrecombinant members is referred to as a parental ditype (PD); one with only recombinant members as a nonparental ditype (NPD); and a tetrad with two recombinant and two nonrecombinant members as a tetratype (TT) (Perkins 1953). If two genetic loci are on different chromosomes, and thus assort independently, the frequency of tetratype (crossover products) versus parental or nonparental assortment ditype (noncrossover products) depends on 10 the frequency of crossover between each of the two loci and their respective centromeres.

Tetratype tetrads arise only when a crossover has occurred between a marker in question and its centromere. Thus, to identify genes that are closely linked to the centromere, markers are examined in a pair-wise fashion until the TT frequency approaches zero. The genetic distance (in centimorgans, cM) between the markers and their respective centromeres is defined 15 by the function  $[(1/2)TT]/100$  (Mortimer *et al.*, 1981). Because positional information obtained by tetrad analysis is a representation of physical distance between two points, as one approaches the centromere the chance of a recombination event declines.

Tetrad analysis has been used to genetically track centromeres in yeasts and other fungi in which products of a single meioses can be collected. The budding yeast *Saccharomyces* 20 *cerevisiae* lacks mitotic condensation and thus cytogenetics (Hegemann *et al.*, 1993), yet due to tetrad analysis, has served as the vehicle of discovery for centromere function. Meiosis is followed by the generation of four spores held within an ascus and these can be directly assayed for gene segregation.

The recessive *qrt1* mutation makes it possible to perform tetrad analysis in *Arabidopsis* 25 by causing the four products of meiosis to remain attached (Preuss *et al.*, 1994; and Smythe 1994; both incorporated herein by reference). As previously shown, within each tetrad, genetic loci segregate in a 2:2 ratio (FIG. 6). Individual tetrads can be manipulated onto flowers with a fine brush (at a rate of 20 tetrads per hour), and in 30% of such crosses, four viable seeds can be obtained (Preuss *et al.*, 1994).

Mapping centromeres with high precision requires a dense genetic map, and although the current *Arabidopsis* map contains many visible markers, it would be laborious to cross each into the *gr1* background. Alternatively, hundreds of DNA polymorphisms can be introduced simultaneously by crossing two different strains, both containing the *gr1* mutation. A dense 5 RFLP map (Chang *et al.*, 1988) and PCR-based maps (Konieczny *et al.*, 1993; Bell *et al.*, 1994) have been generated in *Arabidopsis* from crosses of the Landsberg and Columbia strains (*Arabidopsis* map and genetic marker data is available from the internet at <http://genome-www.stanford.edu/Arabidopsis> and [http://cbil.humgen.upenn.edu/atgc/sslp\\_info/sslp.html](http://cbil.humgen.upenn.edu/atgc/sslp_info/sslp.html)). These strains differ by 1% at the DNA sequence level and have colinear genetic maps (Chang *et* 10 *al.*, 1988; Koornneef, 1987).

Centromere mapping with tetrad analysis requires simultaneous analysis of two markers, one of which must be centromere-linked (FIG. 1). To identify these centromere-linked markers, markers distributed across all 5 chromosomes were scored and compared in a pairwise fashion.

15 Initially, genetic markers that can be scored by PCR analysis were tested (Konieczny *et al.*, 1993; Bell *et al.*, 1994). Such markers are sufficiently dense to map any locus to  $\pm$  10 cM, and as additional PCR-detectable polymorphisms are identified they are incorporated into the analyses. As higher resolution mapping becomes necessary, selected RFLPs, from the existing set of over 350 markers will be scored. In addition, as described in FIG. 5, new CAPS and 20 SSLP markers useful for mapping the centromere can be readily identified.

## II. Tetrad Sets

To date, progeny plants from 388 isolated tetrad seed sets have been germinated and leaf tissue collected and stored from each of the tetrad progeny plants. The leaf tissue from individual plants was used to make DNA for PCR based marker analysis. The plants also were 25 allowed to self-fertilize and the seed they produced was collected (a list of the seed stock of informative individuals used for tetrad analysis is given in FIG. 4). From each of these individual seed sets, seedlings can be germinated and their tissues utilized for making genomic DNA. Tissue pooled from multiple seedlings is useful for making Southern genomic DNA blots for the analysis of restriction fragment length polymorphisms (RFLPs). Informative 30 plants which have been used for tetrad analysis are given in FIG. 4.

### III. Mapping Strategy

The initial mapping of the centromeres was completed with selected PCR™ based markers (such as CAPS and SSLPs) (FIG. 5). These markers were selected from the large number of markers available for *Arabidopsis* and are distributed across the genetic map. Due to 5 the fact that analysis of RFLP markers is labor intensive, those markers that could be analyzed with PCR™-based markers were surveyed first. Only a subset of the original 388 tetrads have provided crossovers close enough to their centromeres as to require the use of RFLP marker segregation analysis. For those tetrads in which existing markers do not identify the region of crossover between centromere and marker, new markers may be developed. This may be 10 accomplished by screening cloned sequences in regions of interest for new potential markers, i.e., RFLPs, CAPS, SSLPs, and the like. The new markers may then be used to generate new data with the relevant tetrads. In addition to the markers in centromeric regions, the segregation of two markers per chromosome that are both, far from, and flank the centromeric regions were 15 assessed. Data from these markers will often be tetratype and thus identify whether a Landsberg to Columbia crossover or a Columbia to Landsberg crossover is visible as the centromere is approached.

Codominant cleaved amplified polymorphic sequences (CAPS) were amplified by PCR™. The PCR™ products were observed on agarose gels prior to cleavage by restriction enzymes. Once cleaved, the products were scored for marker segregation on agarose or 20 acrylamide gels. Simple sequence length polymorphisms (SSLPs) also were amplified by PCR™. The PCR™ products were directly scored by polyacrylamide gel electrophoresis (PAGE), or on agarose gels.

### IV. High Resolution Positioning of Centromeres on the Genetic Map

Southern genomic DNA blots in combination with RFLP analysis may be used to map 25 centromeres with a high degree of resolution. The stored seedling tissue provides the necessary amount of DNA for analysis of the restriction fragments. Southern blots are hybridized to probes labeled by radioactive or non-radioactive methods.

It may, in many cases, be desired to identify new polymorphic DNA markers which are closely linked to the target region. In some cases this can be readily done. For example when 30 comparing Landsberg and Columbia DNA, a polymorphic *Sau3A* site can be found for about

every 8 to 20 kB surveyed. Subtractive methods are available for identifying such polymorphisms (Rosenberg *et al.*, 1994), and these subtractions may be performed using DNA from selected, centromeric YAC clones. Screens for RFLP markers potentially linked to centromeres also can be performed using DNA fragments from a centromere-linked YAC clone 5 to probe blots of Landsberg and Columbia genomic DNA that has been digested with a panel of restriction enzymes.

#### V. Isolation of Centromere Containing DNA Fragments

Using the markers flanking each centromere (see FIG. 3) it is possible to purify a contiguous DNA fragment that contains both flanking markers and the centromere encoded 10 between those markers. In order to carry this out, very large DNA fragments up to the size of an entire chromosome are prepared by embedding *Arabidopsis* tissues in agarose using, for example, the method described by Copenhaver *et al.*, (1995). These large pieces of DNA can be digested in the agarose with any restriction enzyme. Those restriction enzymes which will be particularly useful for isolating intact centromeres include enzymes which yield very large 15 DNA fragments. Such restriction enzymes include those with specificities greater than six base pairs such as, for example, Asc I, Bae I, BbvC I, Fse I, Not I, Pac I, Pme I, PpuM I, Rsr II, SanD I, Sap I, SexA I, Sfi I, Sgf I, SgrA I, Sbf I, Srf I, Sse8387 I, Sse8647 I, Swa, UbaD I, and UbaE I, or any other enzyme that cuts at a low frequency within the *Arabidopsis* genome, and specifically within the centromeric region. Alternatively, a partial digest with a more frequent 20 cutting restriction enzyme could be used.

The large DNA fragments produced by digestion with restriction enzymes are then separated by size using pulsed-field gel electrophoresis (PFGE) (Schwartz *et al.*, 1982). Specifically, Contour-clamped Homogeneous Electric Field (CHEF) electrophoresis (a variety of PFGE) can be used to separate DNA molecules as large as 10 Mb (Chu *et al.*, 1985). Large 25 DNA fragments resolved on CHEF gels can then be analyzed using standard Southern hybridization techniques to identify and measure the size of those fragments which contain both centromere flanking markers and therefore, the centromere. After determining the size of the centromere containing fragment by comparison with known size standards, the region from the gel that contains the centromere fragment can be cut out of a duplicate gel. This centromeric 30 DNA can then be analyzed, sequenced, and used in a variety of applications, as described below, including the construction of plant artificial chromosomes (PLACs). As indicated in

detail below, PLACs can be constructed by attaching telomeres and selectable markers to the centromere fragment cut from the agarose gel using standard techniques which allow DNA ligation within the gel slice. Plant cells can then be transformed with this hybrid DNA molecule using the techniques described herein below.

5      **VI. PLAC Constructs**

In light of the instant disclosure it will be possible for those of ordinary skill in the art to construct the artificial chromosomes described herein. Useful construction methods are well-known (see, for example, Maniatis *et al.*, 1982). As constructed, the PLAC will preferably include at least an autonomous replication sequence (ARS) functional in plants, a centromere 10 functional in plants, and telomeres which are functional in plants.

In addition to the basic elements, positive and negative selectable plant markers (e.g., antibiotic or herbicide resistance genes), and a cloning site for insertion of foreign DNA will 15 preferably be included. In addition, a visible marker, such as green fluorescent protein, also may be desirable. In order to propagate the vectors in *E. coli*, it is necessary to convert the linear molecule into a circle by addition of a stuffer fragment between the telomeres. Inclusion 20 of an *E. coli* plasmid replication origin and selectable marker also may be preferred. It also may be desirable to include *Agrobacterium* sequences to improve replication and transfer to plant cells. Exemplary artificial chromosome constructs are given in FIGs. 7A-7H, although it will be apparent to those in skill art that many changes may be made in the order and types of elements present in these constructs and still obtain a functional artificial chromosome within the scope of the instant invention.

Artificial plant chromosomes which replicate in yeast also may be constructed to take advantage of the large insert capacity and stability of repetitive DNA inserts afforded by this system (see Burke *et al.*, 1987). In this case, yeast ARS and CEN sequences are added to the 25 vector. The artificial chromosome is maintained in yeast as a circular molecule using a stuffer fragment to separate the telomeres.

A fragment of DNA, from any source whatsoever, may be purified and inserted into an artificial plant chromosome at any appropriate restriction endonuclease cleavage site. The DNA segment usually will include various regulatory signals for the expression of proteins

encoded by the fragment. Alternatively, regulatory signals resident in the artificial chromosome may be utilized.

The techniques and procedures required to accomplish insertion are well-known in the art (see Maniatis *et al.*, 1982). Typically, this is accomplished by incubating a circular plasmid or a linear DNA fragment in the presence of a restriction endonuclease such that the restriction endonuclease cleaves the DNA molecule. Endonucleases preferentially break the internal phosphodiester bonds of polynucleotide chains. They may be relatively unspecific, cutting polynucleotide bonds regardless of the surrounding nucleotide sequence. However, the endonucleases which cleave only a specific nucleotide sequence are called restriction enzymes.

5      10      15      20      25

Restriction endonucleases generally internally cleave DNA molecules at specific recognition sites, making breaks within "recognition" sequences that in many, but not all, cases exhibit two-fold symmetry around a given point. Such enzymes typically create double-stranded breaks.

Many of these enzymes make a staggered cleavage, yielding DNA fragments with protruding single-stranded 5' or 3' termini. Such ends are said to be "sticky" or "cohesive" because they will hydrogen bond to complementary 3' or 5' ends. As a result, the end of any DNA fragment produced by an enzyme, such as *EcoRI*, can anneal with any other fragment produced by that enzyme. This properly allows splicing of foreign genes into plasmids, for example. Some restriction endonucleases that may be particularly useful with the current invention include *HindIII*, *PstI*, *EcoRI*, and *BamHI*.

20      25

Some endonucleases create fragments that have blunt ends, that is, that lack any protruding single strands. An alternative way to create blunt ends is to use a restriction enzyme that leaves overhangs, but to fill in the overhangs with a polymerase, such as klenow, thereby resulting in blunt ends. When DNA has been cleaved with restriction enzymes that cut across both strands at the same position, blunt end ligation can be used to join the fragments directly together. The advantage of this technique is that any pair of ends may be joined together, irrespective of sequence.

Those nucleases that preferentially break off terminal nucleotides are referred to as exonucleases. For example, small deletions can be produced in any DNA molecule by treatment with an exonuclease which starts from each 3' end of the DNA and chews away single 30 strands in a 3' to 5' direction, creating a population of DNA molecules with single-stranded

fragments at each end, some containing terminal nucleotides. Similarly, exonucleases that digest DNA from the 5' end or enzymes that remove nucleotides from both strands have often been used. Some exonucleases which may be particularly useful in the present invention include *Bal31*, *SI*, and *ExoIII*. These nucleolytic reactions can be controlled by varying the 5 time of incubation, the temperature, and the enzyme concentration needed to make deletions. Phosphatases and kinases also may be used to control which fragments have ends which can be joined. Examples of useful phosphatases include shrimp alkaline phosphatase and calf intestinal alkaline phosphatase. An example of a useful kinase is T4 polynucleotide kinase.

Once the source DNA sequences and vector sequences have been cleaved and modified 10 to generate appropriate ends they are incubated together with enzymes capable of mediating the ligation of the two DNA molecules. Particularly useful enzymes for this purpose include T4 ligase, *E. coli* ligase, or other similar enzymes. The action of these enzymes results in the sealing of the linear DNA to produce a larger DNA molecule containing the desired fragment (see, for example, U.S. Pat. Nos. 4,237,224; 4,264,731; 4,273,875; 4,322,499 and 4,336,336, 15 which are specifically incorporated herein by reference).

It is to be understood that the termini of the linearized plasmid and the termini of the DNA fragment being inserted must be complementary or blunt in order for the ligation reaction to be successful. Suitable complementarity can be achieved by choosing appropriate restriction endonucleases (*i.e.*, if the fragment is produced by the same restriction endonuclease or one that 20 generates the same overhang as that used to linearize the plasmid, then the termini of both molecules will be complementary). As discussed previously, in a preferred embodiment, at least two classes of the vectors used in the present invention are adapted to receive the foreign oligonucleotide fragments in only one orientation. After joining the DNA segment to the vector, the resulting hybrid DNA can then be selected from among the large population of 25 clones or libraries.

A method useful for the molecular cloning of DNA sequences includes *in vitro* joining of DNA segments, fragmented from a source high molecular weight genomic DNA, to vector DNA molecules capable of independent replication. The cloning vector may include plasmid DNA (see Cohen *et al.*, 1973), phage DNA (see Thomas *et al.*, 1974), SV40 DNA (see 30 Nussbaum *et al.*, 1976), yeast DNA, *E. coli* DNA and most significantly, plant DNA.

A variety of processes are known which may be utilized to effect transformation; *i.e.*, the inserting of a heterologous DNA sequences into a host cell, whereby the host becomes capable of efficient expression of the inserted sequences.

#### VII. Definitions

5 By "transformation" or "transfection" is meant the acquisition in cells of new DNA sequences through incorporation of added DNA. This is the process by which naked DNA, DNA coated with protein, or whole artificial chromosomes are introduced into a cell, resulting in a heritable change.

10 By "gene" is meant a DNA sequence that contains information for construction of a polypeptide or protein, and includes 5' and 3' ends. This also includes genes which encode only RNA products such as tRNA or rRNA genes.

15 As used herein, "eukaryote" refers to living organisms whose cells contain nuclei. A eukaryote may be distinguished from a "prokaryote" which is an organism which lacks nuclei. Prokaryotes and eukaryotes differ fundamentally in the way their genetic information is organized, as well as their patterns of RNA and protein synthesis.

By the term "lower eukaryote" is meant a eukaryote characterized by a comparatively simple physiology and composition, and most often unicellularity. Examples of lower eukaryotes include flagellates, ciliates, and yeast.

20 By contrast, the term "higher eukaryote" means a multicellular eukaryote, typically characterized by its greater complex physiological mechanisms and relatively large size. Generally, complex organisms such as plants and animals are included in this category. Preferred higher eukaryotes to be transformed by the present invention include, for example, monocot and dicot angiosperm species, gymnosperm species, fern species, plant tissue culture cells of these species, animal cells and algal cells. It will of course be understood that 25 prokaryotes and eukaryotes alike may be transformed by the methods of this invention.

As used herein, the term "plant" includes plant cells, plant protoplasts, plant calli, and the like, as well as whole plants regenerated therefrom.

As used herein, "heterologous gene" or "foreign gene" is a structural gene that is foreign, *i.e.*, originating from a donor different from the host or a chemically synthesized gene,

and can include a donor of a different species from the host. The heterologous gene codes for a polypeptide or RNA ordinarily not produced by the organism susceptible to transformation by the expression vehicle. Another type of "heterologous gene" is an altered gene from the host itself, or an unaltered gene which is present in one or more extra copies. One example of such 5 an altered gene useful in the present invention is a mutant gene which encodes a herbicide-resistant form of a normally occurring enzyme.

By "host" is meant any organism that is the recipient of a replicable plasmid, or expression vector comprising an artificial chromosome. Ideally, host strains used for cloning experiments should be free of any restriction enzyme activity that might degrade the foreign 10 DNA used. Preferred examples of host cells for cloning, useful in the present invention, are bacteria such as *Escherichia coli*, *Bacillus subtilis*, *Pseudomonas*, *Streptomyces*, *Salmonella*, and yeast cells such as *S. cerevisiae*. Host cells which can be targeted for expression of an artificial chromosome may be plant cells of any source and specifically include *Arabidopsis*, maize, rice, sugarcane, sorghum, barley, soybeans, tobacco, wheat, tomato, potato, citrus, or 15 any other agronomically or scientifically important species.

By "expression" is meant the process by which a structural gene produces an RNA molecule, typically termed messenger RNA (mRNA). The mRNA is typically, but not always, translated into polypeptide(s).

By "linker" it is meant a DNA molecule, generally up to 50 or 60 nucleotides long and 20 synthesized chemically, or cloned from other vectors. In a preferred embodiment, this fragment contains one, or preferably more than one, restriction enzyme site for a blunt-cutting enzyme and a staggered-cutting enzyme, such as *Bam*HI. One end of the linker fragment is adapted to be ligatable to one end of the linear molecule and the other end is adapted to be ligatable to the other end of the linear molecule.

25 As used herein, a "library" is a pool of random DNA fragments which are cloned. In principle, any gene can be isolated by screening the library with a specific hybridization probe (see, for example, Young *et al.*, 1977). Each library may contain the DNA of a given organism inserted as discrete restriction enzyme-generated fragments or as randomly sheered fragments into many thousands of plasmid vectors. For purposes of the present invention, *E. coli*, yeast,

and *Salmonella* plasmids are particularly useful when the genome inserts come from other organisms.

By "hybridization" is meant the pairing of complementary RNA and DNA strands to produce an RNA-DNA hybrid, or alternatively, the pairing of two DNA single strands from 5 genetically different or the same sources to produce a double stranded DNA molecule.

The term "plasmid" or "cloning vector" as used herein refers to a closed covalently circular extrachromosomal DNA or linear DNA which is able to autonomously replicate in a host cell and which is normally nonessential to the survival of the cell. A wide variety of plasmids and other vectors are known and commonly used in the art (see, for example, Cohen 10 *et al.*, U.S. Pat. No. 4,468,464, which discloses examples of DNA plasmids, and which is specifically incorporated herein by reference).

As used herein, a "probe" is any biochemical reagent (usually tagged in some way for ease of identification), used to identify or isolate a gene, a gene product, a DNA segment or a protein.

15 By "PLAC" it is meant a plant artificial chromosome of the current invention, as specifically disclosed herein.

A "selectable marker" is a gene whose presence results in a clear phenotype, and most often a growth advantage for cells that contain the marker. This growth advantage may be present under standard conditions, altered conditions such as elevated temperature, or in the 20 presence of certain chemicals such as herbicides or antibiotics. Use of selectable markers is described, for example, in Broach *et al.* (1979). Examples of selectable markers include the thymidine kinase gene, the cellular adenine-phosphoribosyltransferase gene and the dihydrylicolate reductase gene, hygromycin phosphotransferase genes, the bar gene and neomycin phosphotransferase genes, among others. Preferred selectable markers in the present 25 invention include genes whose expression confer antibiotic or herbicide resistance to the host cell, sufficient to enable the maintenance of a vector within the host cell, and which facilitate the manipulation of the plasmid into new host cells. Of particular interest in the present invention are proteins conferring cellular resistance to ampicillin, chloramphenicol, tetracycline, G-418, bialaphos, and glyphosate for example.

By "PLAC-encoded protein" it is meant a polypeptide which is encoded by a sequence of a PLAC of the current invention. This includes those proteins encoded by functional sequences of the PLAC, such as selectable markers, telomeres, etc., as well as those proteins encoded by heterologous genes of the PLAC.

5 By "PLAC-associated protein" it is meant a protein encoded by a sequence of the PLAC or a protein which is encoded by host DNA and binds with relatively high affinity to the centromeres of the current invention.

**VIII. Arabidopsis thaliana Centromeric DNA Segments as Hybridization Probes and Primers**

10 In addition to their use in the construction of PLACs, the centromeric regions disclosed herein also have a variety of other uses. For example, they also have utility as probes or primers in nucleic acid hybridization embodiments. As such, it is contemplated that nucleic acid segments that comprise a sequence region that consists of at least a 14 nucleotide long contiguous sequence that has the same sequence as, or is complementary to, a 14 nucleotide long contiguous DNA segment of a centromere of the current invention will find particular utility. Longer contiguous identical or complementary sequences, e.g., those of about 20, 30, 15 40, 50, 100, 200, 500, 1000, 2000, 5000 bp, etc. including all intermediate lengths and up to and including the full-length sequence of a centromere of the current invention also will be of use in certain embodiments.

20 The ability of such nucleic acid probes to specifically hybridize to centromeric sequences will enable them to be of use in detecting the presence of similar, partially complementary sequences from other plants or animals. However, other uses are envisioned, including the use of the centromeres for the preparation of mutant species primers, or primers for use in preparing other genetic constructions.

25 Nucleic acid molecules having sequence regions consisting of contiguous nucleotide stretches of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or even of 101-200 nucleotides or so, identical or complementary to DNA sequences of a centromere of the current invention, are particularly contemplated as hybridization probes for use in, e.g., Southern and Northern blotting. Smaller fragments will

generally find use in hybridization embodiments, wherein the length of the contiguous complementary region may be varied, such as between about 10-14 and about 100 or 200 nucleotides, but larger contiguous complementarity stretches also may be used, according to the length complementary sequences one wishes to detect.

5       Of course, fragments may also be obtained by other techniques such as, *e.g.*, by mechanical shearing or by restriction enzyme digestion. Small nucleic acid segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such 10      as the PCR™ technology of U. S. Patents 4,683,195 and 4,683,202 (each incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular biology.

15      Accordingly, the centromeres of the current invention may be used for their ability to selectively form duplex molecules with complementary stretches of DNA fragments. Depending on the application envisioned, one will desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, *e.g.*, one will select relatively low salt and/or high temperature 20      conditions, such as provided by about 0.02 M to about 0.15 M NaCl at temperatures of about 50°C to about 70°C. Such selective conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating centromeric DNA segments. Detection of DNA segments via hybridization is well-known to those of skill in the art, and the teachings of U. S. Patents 4,965,188 and 5,176,995 (each 25      specifically incorporated herein by reference in its entirety) are exemplary of the methods of hybridization analyses. Teachings such as those found in the texts of Maloy *et al.*, 1991; Segal 1976; Prokop 1991; and Kuby 1994, are particularly relevant.

30      Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template or where one seeks to isolate centromere function-encoding sequences from related species, functional equivalents, or the like, less stringent hybridization conditions will typically be needed in order to allow

formation of the heteroduplex. In these circumstances, one may desire to employ conditions such as about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that 5 conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as increased temperature or decreased salt. Thus, hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

In certain embodiments, it will be advantageous to employ nucleic acid sequences of the 10 present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of giving a detectable signal. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as urease, alkaline phosphatase or peroxidase, instead 15 of radioactive or other environmentally undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known that can be employed to provide a means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

In general, it is envisioned that the hybridization probes described herein will be useful 20 both as reagents in solution hybridization as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to specific hybridization with selected probes under desired conditions. The selected 25 conditions will depend on the particular circumstances based on the particular criteria required (depending, for example, on the G+C content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, *etc.*). Following washing of the hybridized surface so as to remove nonspecifically bound probe molecules, specific hybridization is detected, or even quantitated, by means of the label.

**IX. PLAC-Associated Protein Specific Antibody Compositions and Methods of Making**

In particular embodiments, the inventors contemplate the use of antibodies, either monoclonal or polyclonal which bind to PLAC-associated proteins of the current invention.

5 Such PLAC-associated proteins include proteins which are coded by the sequences of the PLAC, as well as proteins encoded by host DNA that bind to the centromeres of the current invention. It is specifically contemplated that these PLAC-associated protein specific antibodies would allow for the further isolation and characterization of the PLAC-associated proteins. For example, proteins may be isolated which are encoded by the centromeres.

10 Recombinant production of such proteins provides a source of antigen for production of antibodies.

Alternatively, the centromere may be used as a ligand to isolate, using affinity methods, centromere binding proteins. Once isolated, these protein can be used as antigens for the production polyclonal and monoclonal antibodies. A variation on this technique has been

15 demonstrated by Rattner (1991), by cloning of centromere-associated proteins through the use of antibodies which bind in the vicinity of the centromere.

Means for preparing and characterizing antibodies are well known in the art (see, *e.g.*, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, 1988; incorporated herein by reference). The methods for generating monoclonal antibodies (mAbs) generally begin

20 along the same lines as those for preparing polyclonal antibodies. Briefly, a polyclonal antibody is prepared by immunizing an animal with an immunogenic composition in accordance with the present invention and collecting antisera from that immunized animal. A wide range of animal species can be used for the production of antisera. Typically the animal used for production of antisera is a rabbit, a mouse, a rat, a hamster, a guinea pig or a goat. A

25 rabbit is a preferred choice for production of polyclonal antibodies because of the ease of handling, maintenance and relatively large blood volume.

As is well known in the art, a given composition may vary in its immunogenicity. It is often necessary therefore to boost the host immune system, as may be achieved by coupling a peptide or polypeptide immunogen to a carrier. Exemplary and preferred carriers are keyhole limpet hemocyanin (KLH) and bovine serum albumin (BSA). Other albumins such as ovalbumin, mouse serum albumin or rabbit serum albumin also can be used as carriers. Means

for conjugating a polypeptide to a carrier protein are well known in the art and include glutaraldehyde, *m*-maleimidobencoyl-N-hydroxysuccinimide ester, carbodimide and bis-biazotized benzidine.

As is also well known in the art, the immunogenicity of a particular immunogen 5 composition can be enhanced by the use of non-specific stimulators of the immune response, known as adjuvants. Exemplary and preferred adjuvants include complete Freund's adjuvant (a non-specific stimulator of the immune response containing killed *Mycobacterium tuberculosis*), incomplete Freund's adjuvants and aluminum hydroxide adjuvant.

The amount of immunogen composition used in the production of polyclonal antibodies 10 varies upon the nature of the immunogen as well as the animal used for immunization. A variety of routes can be used to administer the immunogen (subcutaneous, intramuscular, intradermal, intravenous and intraperitoneal). The production of polyclonal antibodies may be monitored by sampling blood of the immunized animal at various points following immunization. A second, booster, injection also may be given. The process of boosting and 15 titering is repeated until a suitable titer is achieved. When a desired level of immunogenicity is obtained, the immunized animal can be bled and the serum isolated and stored, and/or the animal can be used to generate mAbs.

Monoclonal antibodies may be readily prepared through use of well-known techniques, such as those exemplified in U. S. Patent 4,196,265, incorporated herein by reference. 20 Typically, this technique involves immunizing a suitable animal with a selected immunogen composition, *e.g.*, a purified or partially purified PLAC-associated protein, polypeptide or peptide. The immunizing composition is administered in a manner effective to stimulate antibody producing cells. Rodents such as mice and rats are preferred animals, however, the use of rabbit, sheep, or frog cells also is possible. The use of rats may provide certain 25 advantages (Goding 1986), but mice are preferred, with the BALB/c mouse being most preferred as this is most routinely used and generally gives a higher percentage of stable fusions.

Following immunization, somatic cells with the potential for producing antibodies, specifically B lymphocytes (B cells), are selected for use in the mAb generating protocol. 30 These cells may be obtained from biopsied spleens, tonsils or lymph nodes, or from a peripheral

blood sample. Spleen cells and peripheral blood cells are preferred, the former because they are a rich source of antibody-producing cells that are in the dividing plasmablast stage, and the latter because peripheral blood is easily accessible. Often, a panel of animals will have been immunized and the spleen of animal with the highest antibody titer will be removed and the 5 spleen lymphocytes obtained by homogenizing the spleen with a syringe. Typically, a spleen from an immunized mouse contains approximately  $5 \times 10^7$  to  $2 \times 10^8$  lymphocytes.

The antibody-producing B lymphocytes from the immunized animal are then fused with cells of an immortal myeloma cell, generally one of the same species as the animal that was immunized. Myeloma cell lines suited for use in hybridoma-producing fusion procedures 10 preferably are non-antibody-producing, have high fusion efficiency, and enzyme deficiencies that render them incapable of growing in certain selective media which support the growth of the desired fused cells (hybridomas).

Any one of a number of myeloma cells may be used, as are known to those of skill in the art (Goding 1986; Campbell 1984). For example, where the immunized animal is a mouse, 15 one may use P3-X63/Ag8, X63-Ag8.653, NS1/1.Ag 4 1, Sp210-Ag14, FO, NSO/U, MPC-11, MPC11-X45-GTG 1.7 and S194/5XX0 Bul; for rats, one may use R210.RCY3, Y3-Ag 1.2.3, IR983F and 4B210; and U-266, GM1500-GRG2, LICR-LON-HMy2 and UC729-6 are all useful in connection with human cell fusions.

One preferred murine myeloma cell is the NS-1 myeloma cell line (also termed P3-NS-20 1-Ag4-1), which is readily available from the NIGMS Human Genetic Mutant Cell Repository by requesting cell line repository number GM3573. Another mouse myeloma cell line that may be used is the 8-azaguanine-resistant mouse murine myeloma SP2/0 non-producer cell line.

Methods for generating hybrids of antibody-producing spleen or lymph node cells and myeloma cells usually comprise mixing somatic cells with myeloma cells in a 2:1 ratio, though 25 the ratio may vary from about 20:1 to about 1:1, respectively, in the presence of an agent or agents (chemical or electrical) that promote the fusion of cell membranes. Fusion methods using Sendai virus have been described (Kohler *et al.*, 1975; 1976), and those using polyethylene glycol (PEG), such as 37% (v/v) PEG, (Gefter *et al.*, 1977). The use of electrically induced fusion methods also is appropriate (Goding 1986).

Fusion procedures usually produce viable hybrids at low frequencies, about  $1 \times 10^{-6}$  to  $1 \times 10^{-8}$ . However, this does not pose a problem, as the viable, fused hybrids are differentiated from the parental, unfused cells (particularly the unfused myeloma cells that would normally continue to divide indefinitely) by culturing in a selective medium. The selective medium is 5 generally one that contains an agent that blocks the *de novo* synthesis of nucleotides in the tissue culture media. Exemplary and preferred agents are aminopterin, methotrexate, and azaserine. Aminopterin and methotrexate block *de novo* synthesis of both purines and pyrimidines, whereas azaserine blocks only purine synthesis. Where aminopterin or methotrexate is used, the media is supplemented with hypoxanthine and thymidine as a source 10 of nucleotides (HAT medium). Where azaserine is used, the media is supplemented with hypoxanthine.

The preferred selection medium is HAT. Only cells capable of operating nucleotide salvage pathways are able to survive in HAT medium. The myeloma cells are defective in key enzymes of the salvage pathway, *e.g.*, hypoxanthine phosphoribosyl transferase (HPRT), and 15 they cannot survive. The B-cells can operate this pathway, but they have a limited life span in culture and generally die within about two weeks. Therefore, the only cells that can survive in the selective media are those hybrids formed from myeloma and B-cells.

This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone 20 dilution in microtiter plates, followed by testing the individual clonal supernatants (after about two to three weeks) for the desired reactivity. The assay should be sensitive, simple and rapid, such as radioimmunoassays, enzyme immunoassays, cytotoxicity assays, plaque assays, dot immunobinding assays, and the like.

The selected hybridomas would then be serially diluted and cloned into individual 25 antibody-producing cell lines, which clones can then be propagated indefinitely to provide mAbs. The cell lines may be exploited for mAb production in two basic ways. A sample of the hybridoma can be injected (often into the peritoneal cavity) into a histocompatible animal of the type that was used to provide the somatic and myeloma cells for the original fusion. The injected animal develops tumors secreting the specific monoclonal antibody produced by the 30 fused cell hybrid. The body fluids of the animal, such as serum or ascites fluid, can then be tapped to provide mAbs in high concentration. The individual cell lines also could be cultured

*in vitro*, where the mAbs are naturally secreted into the culture medium from which they can be readily obtained in high concentrations. mAbs produced by either means may be further purified, if desired, using filtration, centrifugation and various chromatographic methods such as HPLC or affinity chromatography.

5    X.    ELISAs and Immunoprecipitation

ELISAs may be used in conjunction with the invention. Particularly, it is contemplated that ELISAs will find use in assays of PLAC gene expression. In an ELISA assay, proteins or peptides comprising PLAC-encoded protein antigen sequences are immobilized onto a selected surface, preferably a surface exhibiting a protein affinity such as the wells of a polystyrene 10 microtiter plate. After washing to remove incompletely adsorbed material, it is desirable to bind or coat the assay plate wells with a nonspecific protein that is known to be antigenically neutral with regard to the test antisera such as bovine serum albumin (BSA), casein or solutions 15 of milk powder. This allows for blocking of nonspecific adsorption sites on the immobilizing surface and thus reduces the background caused by nonspecific binding of antisera onto the surface.

After binding of antigenic material to the well, coating with a non-reactive material to reduce background, and washing to remove unbound material, the immobilizing surface is contacted with the antisera or clinical or biological extract to be tested in a manner conducive to immune complex (antigen/antibody) formation. Such conditions preferably include diluting the 20 antisera with diluents such as BSA, bovine gamma globulin (BGG) and phosphate buffered saline (PBS)/Tween®. These added agents also tend to assist in the reduction of nonspecific background. The layered antisera is then allowed to incubate for from about 2 to about 4 hours, at temperatures preferably on the order of about 25° to about 27°C. Following incubation, the 25 antisera-contacted surface is washed so as to remove non-immunocomplexed material. A preferred washing procedure includes washing with a solution such as PBS/Tween®, or borate buffer.

Following formation of specific immunocomplexes between the test sample and the bound antigen, and subsequent washing, the occurrence and even amount of immunocomplex formation may be determined by subjecting same to a second antibody having specificity for 30 the first. To provide a detecting means, the second antibody will preferably have an associated

enzyme that will generate color or light development upon incubating with an appropriate chromogenic substrate. Thus, for example, one will desire to contact and incubate the antisera-bound surface with a urease or peroxidase-conjugated anti-human IgG for a period of time and under conditions which favor the development of immunocomplex formation (e.g., incubation 5 for 2 hours at room temperature in a PBS-containing solution).

After incubation with the second enzyme-tagged antibody, and subsequent to washing to remove unbound material, the amount of label is quantified by incubation with a chromogenic substrate such as urea and bromocresol purple or 2,2'-azino-di-(3-ethylbenzthiazoline)-6-sulfonic acid (ABTS) and H<sub>2</sub>O<sub>2</sub>, in the case of peroxidase as the enzyme 10 label. Quantitation is then achieved by measuring the degree of color generation, e.g., using a visible spectra spectrophotometer.

#### **XI. Western Blots**

The compositions of the present invention may find use in immunoblot or western blot analysis. The antibodies of the invention may be used as high-affinity primary reagents for the 15 identification of proteins immobilized onto a solid support matrix, such as nitrocellulose, nylon or combinations thereof. In conjunction with immunoprecipitation, followed by gel electrophoresis, these may be used as a single step reagent for use in detecting antigens against which secondary reagents used in the detection of the antigen cause an adverse background. This is especially useful when the antigens studied are immunoglobulins (precluding the use of 20 immunoglobulins binding bacterial cell wall components), the antigens studied cross-react with the detecting agent, or they migrate at the same relative molecular weight as a cross-reacting signal.

Immunologically-based detection methods for use in conjunction with Western blotting include enzymatically-, radiolabel-, or fluorescently-tagged secondary antibodies against the 25 protein moiety are considered to be of particular use in this regard.

## XII. DNA Segments

Further aspects of the present invention concern isolated DNA segments and recombinant vectors encoding a functional *Arabidopsis thaliana* centromere and other sequences for the creation and use of recombinant sequences of a PLAC.

5 The present invention concerns DNA segments, isolatable from *A. thaliana* cells, that are enriched relative to total genomic DNA and are capable of conferring centromere activity to a recombinant molecule when incorporated into the host cell. As used herein, the term centromere activity indicates the ability to confer stable inheritance upon a DNA segment, artificial chromosome or chromosome.

10 As used herein, the term "DNA segment" refers to a DNA molecule that has been purified from total genomic DNA of a particular species. Therefore, a DNA segment encoding centromere function refers to a DNA segment that contains centromere coding sequences yet is isolated away from, or purified free from, total genomic DNA of *A. thaliana*. Included within the term "DNA segment", are DNA segments and smaller fragments of such segments, and also 15 recombinant vectors, including, for example, plasmids, cosmids, phage, viruses, and the like.

Similarly, a DNA segment comprising an isolated or purified centromeric sequence refers to a DNA segment including centromere coding sequences and, in certain aspects, regulatory sequences, isolated substantially away from other naturally occurring genes, protein encoding sequences, or other DNA sequences. In this respect, the term "gene" is used for 20 simplicity to refer to a functional DNA segment, protein, polypeptide or peptide encoding unit. As will be understood by those in the art, this functional term includes both genomic sequences, cDNA sequences and smaller engineered gene segments that may express, or may be adapted to express, proteins, polypeptides or peptides.

"Isolated substantially away from other coding sequences" means that the sequences of 25 interest, in this case centromere function encoding sequences, are included within the genomic DNA clones provided herein. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

In particular embodiments, the invention concerns isolated DNA segments and 30 recombinant vectors incorporating DNA sequences that encode a centromere functional sequence that includes a contiguous sequence from the centromeres of the current invention. In

certain other embodiments, the invention concerns isolated DNA segments and recombinant vectors that include within their sequence a contiguous nucleic acid sequence from an *A. thaliana* centromere. Again, DNA segments that exhibit centromere function activity will be most preferred.

5        The nucleic acid segments of the present invention, regardless of the length of the sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total 10 length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

### **XIII. Biological Functional Equivalents**

Modification and changes may be made in the centromeric DNA segments of the current invention and still obtain a functional molecule with desirable characteristics. The 15 following is a discussion based upon changing the nucleic acids of a centromere to create an equivalent, or even an improved, second-generation molecule.

In particular embodiments of the invention, mutated centromeric sequences are contemplated to be useful for increasing the utility of the centromere. It is specifically contemplated that the function of the centromeres of the current invention may be based upon the secondary structure of the DNA sequences of the centromere and / or the proteins which 20 interact with the centromere. By changing the DNA sequence of the centromere, one may alter the affinity of one or more centromere-associated protein(s) for the centromere and / or the secondary structure of the centromeric sequences, thereby changing the activity of the centromere. Alternatively, changes may be made in the centromeres of the invention which do 25 not effect the activity of the centromere. Changes in the centromeric sequences which reduce the size of the DNA segment needed to confer centromere activity are contemplated to be particularly useful in the current invention, as would changes which increased the fidelity with which the centromere was transmitted during mitosis and meiosis.

#### **XIV. Obtaining *A. thaliana* Centromeric DNA from Yeast and Bacterial Clones**

The *Arabidopsis* physical map consists primarily of YAC clones between 200 kb and 1 Mb in length, and an overlapping contig map with several anchors to the genetic map is available (Hwang *et al.*, 1991) (<http://cbil.humgen.upenn.edu/atgc/ATGCUP.html>). To be 5 certain that an entire centromeric region has been cloned, clones or a series of clones, are identified that hybridize to markers on either side of each centromere. These efforts can be complicated by the presence of repetitive DNA in the centromere, as well as by the potential instability of centromere clones. Thus, identification of large YACs with unique sequences that will serve as useful probes simplifies a chromosome walking strategy. Use of BACs may also 10 be advantageous, as data has suggested that YAC clones may sometimes not span centromeres (Willard, 1997).

Blot hybridization allows comparison of the structure of the clones with that of genomic DNA, and thus determines whether the clones have suffered deletions or rearrangements. The centromeric clones identified are useful for hybridization experiments that can be used to 15 determine whether they share common sequences, whether they localize *in situ* to the cytologically defined centromeric region, and whether they contain repetitive sequences thought to map near *Arabidopsis* centromeres (Richards *et al.*, 1991; Maluszynska *et al.*, 1991).

In a positional cloning approach, one may wish to begin sequencing if the region of interest has been narrowed to a sufficiently small area. The determination of what constitutes a 20 sufficiently small sequence is dependent upon many factors including, but not necessarily limited to, the repetitive DNA content of the region of interest, the size of the region to be sequenced, the ability to obtain higher resolution genetic mapping data, and the relative efficiency of sequencing technology. In many cases, sequencing may be begun when the target region has been narrowed to 40 kB or less. It is estimated that in *Arabidopsis thaliana* 1 cM 25 corresponds to approximately 200 kB (Hwang, 1991; Koornneef, 1987). Such a resolution will on average, be provided by analysis of about 100 tetrads for each centromere, although some centromeres may be mapped with higher resolution due to the fact that each tetrad is simultaneously mapping five centromeres, and thus has five times the usual probability of finding a crossover very close to a locus of interest. Nonetheless, because the amount of DNA 30 per cM can increase in the vicinity of the centromere (Carpenter *et al.*, 1982), examination of

additional tetrad progeny and identification of additional genetic markers may be required to achieve the desired degree of genetic linkage.

#### **XV. Transformed Host Cells and Transgenic Plants**

Methods and compositions for transforming a bacterium, a yeast cell, a plant cell, or an 5 entire plant with one or more artificial chromosomes are further aspects of this disclosure. A transgenic bacterium, yeast cell, plant cell or plant derived from such a transformation process or the progeny and seeds from such a transgenic plant also are further embodiments of the invention.

Means for transforming bacteria and yeast cells are well known in the art. Typically, 10 means of transformation are similar to those well known means used to transform other bacteria or yeast such as *E. coli* or *Saccharomyces cerevisiae*. Methods for DNA transformation of plant cells include *Agrobacterium*-mediated plant transformation, protoplast transformation, gene transfer into pollen, injection into reproductive organs, injection into immature embryos and particle bombardment. Each of these methods has distinct advantages and disadvantages. 15 Thus, one particular method of introducing genes into a particular plant strain may not necessarily be the most effective for another plant strain, but it is well known in the art which methods are useful for a particular plant strain.

There are many methods for introducing transforming DNA segments into cells, but not 20 all are suitable for delivering DNA to plant cells. Suitable methods are believed to include virtually any method by which DNA can be introduced into a cell, such as by *Agrobacterium* infection, direct delivery of DNA such as, for example, by PEG-mediated transformation of protoplasts (Omirulleh *et al.*, 1993), by desiccation/inhibition-mediated DNA uptake, by electroporation, by agitation with silicon carbide fibers, by acceleration of DNA coated particles, *etc.* In certain embodiments, acceleration methods are preferred and include, for 25 example, microprojectile bombardment and the like.

Technology for introduction of DNA into cells is well-known to those of skill in the art. 30 Four general methods for delivering a gene into cells have been described: (1) chemical methods (Graham *et al.*, 1973; Zatloukal *et al.*, 1992); (2) physical methods such as microinjection (Capechi, 1980), electroporation (Wong *et al.*, 1982; Fromm *et al.*, 1985; U. S. Patent No. 5,384,253) and the gene gun (Johnston *et al.*, 1994; Fynan *et al.*, 1993); (3) viral

vectors (Clapp 1993; Lu *et al.*, 1993; Eglitis *et al.*, 1988a; 1988b); and (4) receptor-mediated mechanisms (Curiel *et al.*, 1991; 1992; Wagner *et al.*, 1992).

(i) *Electroporation*

The application of brief, high-voltage electric pulses to a variety of animal and plant cells leads to the formation of nanometer-sized pores in the plasma membrane. DNA is taken directly into the cell cytoplasm either through these pores or as a consequence of the redistribution of membrane components that accompanies closure of the pores. Electroporation can be extremely efficient and can be used both for transient expression of cloned genes and for establishment of cell lines that carry integrated copies of the gene of interest. Electroporation, in contrast to calcium phosphate-mediated transfection and protoplast fusion, frequently gives rise to cell lines that carry one, or at most a few, integrated copies of the foreign DNA.

The introduction of DNA by means of electroporation, is well-known to those of skill in the art. In this method, certain cell wall-degrading enzymes, such as pectin-degrading enzymes, are employed to render the target recipient cells more susceptible to transformation by electroporation than untreated cells. Alternatively, recipient cells are made more susceptible to transformation, by mechanical wounding. To effect transformation by electroporation one may employ either friable tissues such as a suspension culture of cells, or embryogenic callus, or alternatively, one may transform immature embryos or other organized tissues directly. One would partially degrade the cell walls of the chosen cells by exposing them to pectin-degrading enzymes (pectolyases) or mechanically wounding in a controlled manner. Such cells would then be recipient to DNA transfer by electroporation, which may be carried out at this stage, and transformed cells then identified by a suitable selection or screening protocol dependent on the nature of the newly incorporated DNA.

(ii) *Microprojectile Bombardment*

A further advantageous method for delivering transforming DNA segments to plant cells is microprojectile bombardment. In this method, particles may be coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, gold, platinum, and the like.

An advantage of microprojectile bombardment, in addition to it being an effective means of reproducibly stably transforming monocots, is that neither the isolation of protoplasts

(Cristou *et al.*, 1988) nor the susceptibility to *Agrobacterium* infection is required. An illustrative embodiment of a method for delivering DNA into maize cells by acceleration is a Biolistics Particle Delivery System, which can be used to propel particles coated with DNA or cells through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered 5 with plant cells cultured in suspension. The screen disperses the particles so that they are not delivered to the recipient cells in large aggregates. It is believed that a screen intervening between the projectile apparatus and the cells to be bombarded reduces the size of projectiles aggregate and may contribute to a higher frequency of transformation by reducing damage inflicted on the recipient cells by projectiles that are too large.

10 For the bombardment, cells in suspension are preferably concentrated on filters or solid culture medium. Alternatively, immature embryos or other target cells may be arranged on solid culture medium. The cells to be bombarded are positioned at an appropriate distance below the macroprojectile stopping plate. If desired, one or more screens also are positioned between the acceleration device and the cells to be bombarded. Through the use of techniques 15 set forth herein one may obtain up to 1000 or more foci of cells transiently expressing a marker gene. The number of cells in a focus which express the exogenous gene product 48 hours post-bombardment often range from 1 to 10 and average 1 to 3.

In bombardment transformation, one may optimize the prebombardment culturing conditions and the bombardment parameters to yield the maximum numbers of stable 20 transformants. Both the physical and biological parameters for bombardment are important in this technology. Physical factors are those that involve manipulating the DNA/microprojectile precipitate or those that affect the flight and velocity of either the macro- or microprojectiles. Biological factors include all steps involved in manipulation of cells before and immediately 25 after bombardment, the osmotic adjustment of target cells to help alleviate the trauma associated with bombardment, and also the nature of the transforming DNA, such as linearized DNA or intact supercoiled plasmids. It is believed that pre-bombardment manipulations are especially important for successful transformation of immature embryos.

Accordingly, it is contemplated that one may wish to adjust various of the bombardment 30 parameters in small scale studies to fully optimize the conditions. One may particularly wish to adjust physical parameters such as gap distance, flight distance, tissue distance, and helium pressure. One also may minimize the trauma reduction factors (TRFs) by modifying conditions

which influence the physiological state of the recipient cells and which may therefore influence transformation and integration efficiencies. For example, the osmotic state, tissue hydration and the subculture stage or cell cycle of the recipient cells may be adjusted for optimum transformation. The execution of other routine adjustments will be known to those of skill in 5 the art in light of the present disclosure.

(iii) *Agrobacterium*-Mediated Transfer

*Agrobacterium*-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, thereby bypassing the need for regeneration of an intact plant from a protoplast. The use of *Agrobacterium*-mediated plant integrating vectors to introduce DNA into plant cells is well known in the art. See, for example, the methods described (Fraley *et al.*, 1985; Rogers *et al.*, 1987). Using conventional transformation vectors, chromosomal integration is required for stable inheritance 10 of the foreign DNA. However, the vector described herein may be used for transformation with or without integration, as the centromere function required for stable inheritance is encoded within the PLAC. In particular embodiments, transformation events in which the 15 PLAC is not chromosomally integrated may be preferred, in that problems with site-specific variations in expression and insertional mutagenesis may be avoided.

The integration of the Ti-DNA is a relatively precise process resulting in few rearrangements. The region of DNA to be transferred is defined by the border sequences, and 20 intervening DNA is usually inserted into the plant genome as described (Spielmann *et al.*, 1986; Jorgensen *et al.*, 1987). Modern *Agrobacterium* transformation vectors are capable of replication in *E. coli* as well as *Agrobacterium*, allowing for convenient manipulations as described (Klee *et al.*, 1985). Moreover, recent technological advances in vectors for 25 *Agrobacterium*-mediated gene transfer have improved the arrangement of genes and restriction sites in the vectors to facilitate construction of vectors capable of expressing various polypeptide coding genes. The vectors described (Rogers *et al.*, 1987), have convenient multi-linker regions flanked by a promoter and a polyadenylation site for direct expression of inserted 30 polypeptide coding genes and are suitable for present purposes. In addition, *Agrobacterium* containing both armed and disarmed Ti genes can be used for the transformations. In those plant strains where *Agrobacterium*-mediated transformation is efficient, it is the method of choice because of the facile and defined nature of the gene transfer.

Agrobacterium-mediated transformation of leaf disks and other tissues such as cotyledons and hypocotyls appears to be limited to plants that *Agrobacterium* naturally infects. *Agrobacterium*-mediated transformation is most efficient in dicotyledonous plants. Few monocots appear to be natural hosts for *Agrobacterium*, although transgenic plants have been produced in asparagus and more significantly in maize using *Agrobacterium* vectors as described (Bytebier *et al.*, 1987; U.S. Patent No. 5,591,616, specifically incorporated herein by reference). Therefore, commercially important cereal grains such as rice, corn, and wheat must usually be transformed using alternative methods. However, as mentioned above, the transformation of asparagus using *Agrobacterium* also can be achieved (see, for example, 5 Bytebier *et al.*, 1987).

A transgenic plant formed using *Agrobacterium* transformation methods typically contains a single gene on one chromosome. Such transgenic plants can be referred to as being hemizygous for the added gene. A more accurate name for such a plant is an independent segregant, because each transformed plant represents a unique T-DNA integration event.

More preferred is a transgenic plant that is homozygous for the added foreign DNA; *i.e.*, a transgenic plant that contains two copies of a transgene, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) an independent segregant transgenic plant that contains a single added transgene, germinating some of the seed produced and analyzing the resulting plants produced for enhanced activity relative to a control (native, non-transgenic) or an independent segregant transgenic plant.

Even more preferred is a plant in which the PLAC has not been chromosomally integrated. Such a plant may be termed  $2n + x$ , where  $2n$  is the diploid number of chromosomes and where  $x$  is the number of PLACs. Initially, transformants may be  $2n+1$ , *i.e.* having 1 additional PLAC. In this case, it may be desirable to self the plant or to cross the plant with another  $2n + 1$  plant to yield a plant which is  $2n + 2$ . The  $2n + 2$  plant is preferred in that it is expected to pass the PLAC through meiosis to all its offspring.

It is to be understood that two different transgenic plants also can be mated to produce offspring that contain two independently segregating added, exogenous PLACs. Selfing of 30 appropriate progeny can produce plants that are homozygous for both added, exogenous PLACs

that encode a polypeptide of interest. Back-crossing to a parental plant and out-crossing with a non-transgenic plant also are contemplated.

#### **XVI. Other Transformation Methods**

Transformation of plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments (see, *e.g.*, Potrykus *et al.*, 1985; Lorz *et al.*, 1985; Fromm *et al.*, 1986; Uchimiya *et al.*, 1986; Callis *et al.*, 1987; Marcotte *et al.*, 1988).

Application of these systems to different plant strains for the purpose of making transgenic plants depends upon the ability to regenerate that particular plant strain from protoplasts. Illustrative methods for the regeneration of cereals from protoplasts are described (Fujimura *et al.*, 1985; Toriyama *et al.*, 1986; Yamada *et al.*, 1986; Abdullah *et al.*, 1986).

To transform plant strains that cannot be successfully regenerated from protoplasts, other ways to introduce DNA into intact cells or tissues can be utilized. For example, regeneration of cereals from immature embryos or explants can be effected as described (Vasil 1988). In addition, "particle gun" or high-velocity microprojectile technology can be utilized (Vasil 1992).

Using that latter technology, DNA is carried through the cell wall and into the cytoplasm on the surface of small metal particles as described (Klein *et al.*, 1987; Klein *et al.*, 1988; McCabe *et al.*, 1988). The metal particles penetrate through several layers of cells and thus allow the transformation of cells within tissue explants.

#### **XVII. Exogenous Genes for Expression in Plants**

One particularly important advance of the present invention is that it provides methods and compositions for expression of exogenous genes in plant cells. Significantly, the current invention allows for the transformation of plant cells with a PLAC comprising a number of exogenous genes. Such genes often will be genes that direct the expression of a particular protein or polypeptide product, but they also may be non-expressible DNA segments, *e.g.*, transposons such as Ds that do not direct their own transposition. As used herein, an "expressible gene" is any gene that is capable of being transcribed into RNA (*e.g.*, mRNA, antisense RNA, *etc.*) or translated into a protein, expressed as a trait of interest, or the like, *etc.*,

and is not limited to selectable, screenable or non-selectable marker genes. The inventors also contemplate that, where both an expressible gene that is not necessarily a marker gene is employed in combination with a marker gene, one may employ the separate genes on either the same or different DNA segments for transformation. In the latter case, the different vectors are 5 delivered concurrently to recipient cells to maximize cotransformation.

The choice of the particular DNA segments to be delivered to the recipient cells often will depend on the purpose of the transformation. One of the major purposes of transformation of crop plants is to add some commercially desirable, agronomically important traits to the plant. Such traits include, but are not limited to, herbicide resistance or tolerance; insect 10 resistance or tolerance; disease resistance or tolerance (viral, bacterial, fungal, nematode); stress tolerance and/or resistance, as exemplified by resistance or tolerance to drought, heat, chilling, freezing, excessive moisture, salt stress; oxidative stress; increased yields; food content and makeup; physical appearance; male sterility; drydown; standability; prolificacy; starch quantity and quality; oil quantity and quality; protein quality and quantity; amino acid composition; and 15 the like. One may desire to incorporate one or more genes conferring any such desirable trait or traits, such as, for example, a gene or genes encoding herbicide resistance.

In certain embodiments, the present invention contemplates the transformation of a recipient cell with PLACs comprising more than one exogenous gene. As used herein, an "exogenous gene," is a gene not normally found in the host genome in an identical context. By 20 this, it is meant that the gene may be isolated from a different species than that of the host genome, or alternatively, isolated from the host genome but operably linked to one or more regulatory regions which differ from those found in the unaltered, native gene. Two or more exogenous genes also can be supplied in a single transformation event using either distinct transgene-encoding vectors, or using a single vector incorporating two or more gene coding 25 sequences. For example, plasmids bearing the *bar* and *aroA* expression units in either convergent, divergent, or colinear orientation, are considered to be particularly useful. Further preferred combinations are those of an insect resistance gene, such as a Bt gene, along with a protease inhibitor gene such as *pinII*, or the use of *bar* in combination with either of the above genes. Of course, any two or more transgenes of any description, such as those conferring 30 herbicide, insect, disease (viral, bacterial, fungal, nematode) or drought resistance, male

sterility, drydown, standability, prolificacy, starch properties, oil quantity and quality, or those increasing yield or nutritional quality may be employed as desired.

(i) *Herbicide Resistance*

The genes encoding phosphinothricin acetyltransferase (*bar* and *pat*), glyphosate tolerant EPSP synthase genes, the glyphosate degradative enzyme gene *gox* encoding glyphosate oxidoreductase, *deh* (encoding a dehalogenase enzyme that inactivates dalapon), herbicide resistant (e.g., sulfonylurea and imidazolinone) acetolactate synthase, and *bxn* genes (encoding a nitrilase enzyme that degrades bromoxynil) are good examples of herbicide resistant genes for use in transformation. The *bar* and *pat* genes code for an enzyme, phosphinothricin acetyltransferase (PAT), which inactivates the herbicide phosphinothricin and prevents this compound from inhibiting glutamine synthetase enzymes. The enzyme 5-enolpyruvylshikimate 3-phosphate synthase (EPSP Synthase), is normally inhibited by the herbicide N-(phosphonomethyl)glycine (glyphosate). However, genes are known that encode glyphosate-resistant EPSP synthase enzymes. These genes are particularly contemplated for use in plant transformation. The *deh* gene encodes the enzyme dalapon dehalogenase and confers resistance to the herbicide dalapon. The *bxn* gene codes for a specific nitrilase enzyme that converts bromoxynil to a non-herbicidal degradation product.

(ii) *Insect Resistance*

Potential insect resistance genes that can be introduced include *Bacillus thuringiensis* crystal toxin genes or Bt genes (Watrud *et al.*, 1985). Bt genes may provide resistance to lepidopteran or coleopteran pests such as European Corn Borer (ECB). Preferred Bt toxin genes for use in such embodiments include the *CryIA(b)* and *CryIA(c)* genes. Endotoxin genes from other species of *B. thuringiensis* which affect insect growth or development also may be employed in this regard.

It is contemplated that preferred Bt genes for use in the transformation protocols disclosed herein will be those in which the coding sequence has been modified to effect increased expression in plants, and more particularly, in monocot plants. Means for preparing synthetic genes are well known in the art and are disclosed in, for example, U.S. Patent No. 5,500,365 and U.S. Patent Number No. 5,689,052, each of the disclosures of which are specifically incorporated herein by reference in their entirety. Examples of such modified Bt

toxin genes include a synthetic Bt *CryIA(b)* gene (Perlak *et al.*, 1991), and the synthetic *CryIA(c)* gene termed 1800b (PCT Application WO 95/06128). Some examples of other Bt toxin genes known to those of skill in the art are given in Table 1 below.

**Table 1**  
*Bacillus thuringiensis*  $\delta$ -Endotoxin Genes<sup>a</sup>

| New Nomenclature | Old Nomenclature | GenBank Accession |
|------------------|------------------|-------------------|
| Cry1Aa           | CryIA(a)         | M11250            |
| Cry1Ab           | CryIA(b)         | M13898            |
| Cry1Ac           | CryIA(c)         | M11068            |
| Cry1Ad           | CryIA(d)         | M73250            |
| Cry1Ae           | CryIA(e)         | M65252            |
| Cry1Ba           | CryIB            | X06711            |
| Cry1Bb           | ET5              | L32020            |
| Cry1Bc           | PEG5             | Z46442            |
| Cry1Bd           | CryE1            | U70726            |
| Cry1Ca           | CryIC            | X07518            |
| Cry1Cb           | CryIC(b)         | M97880            |
| Cry1Da           | CryID            | X54160            |
| Cry1Db           | PrtB             | Z22511            |
| Cry1Ea           | CryIE            | X53985            |
| Cry1Eb           | CryIE(b)         | M73253            |
| Cry1Fa           | CryIF            | M63897            |
| Cry1Fb           | PrtD             | Z22512            |
| Cry1Ga           | PrtA             | Z22510            |
| Cry1Gb           | CryH2            | U70725            |
| Cry1Ha           | PrtC             | Z22513            |
| Cry1Hb           |                  | U35780            |
| Cry1Ia           | CryV             | X62821            |
| Cry1Ib           | CryV             | U07642            |
| Cry1Ja           | ET4              | L32019            |

Table 1 - Continued

| New Nomenclature | Old Nomenclature | GenBank Accession |
|------------------|------------------|-------------------|
| Cry1Jb           | ET1              | U31527            |
| Cry1K            |                  | U28801            |
| Cry2Aa           | CryIIA           | M31738            |
| Cry2Ab           | CryIIB           | M23724            |
| Cry2Ac           | CryIIC           | X57252            |
| Cry3A            | CryIIIA          | M22472            |
| Cry3Ba           | CryIIIB          | X17123            |
| Cry3Bb           | CryIIIB2         | M89794            |
| Cry3C            | CryIID           | X59797            |
| Cry4A            | CryIVA           | Y00423            |
| Cry4B            | CryIVB           | X07423            |
| Cry5Aa           | CryVA(a)         | L07025            |
| Cry5Ab           | CryVA(b)         | L07026            |
| Cry6A            | CryVIA           | L07022            |
| Cry6B            | CryVIB           | L07024            |
| Cry7Aa           | CryIIIC          | M64478            |
| Cry7Ab           | CryIIICb         | U04367            |
| Cry8A            | CryIIIE          | U04364            |
| Cry8B            | CryIIIG          | U04365            |
| Cry8C            | CryIIIF          | U04366            |
| Cry9A            | CryIG            | X58120            |
| Cry9B            | CryIX            | X75019            |
| Cry9C            | CryIH            | Z37527            |
| Cry10A           | CryIVC           | M12662            |
| Cry11A           | CryIVD           | M31737            |
| Cry11B           | Jeg80            | X86902            |
| Cry12A           | CryVB            | L07027            |
| Cry13A           | CryVC            | L07023            |

Table 1 - Continued

| New Nomenclature | Old Nomenclature | GenBank Accession |
|------------------|------------------|-------------------|
| Cry14A           | CryVD            | U13955            |
| Cry15A           | 34kDa            | M76442            |
| Cry16A           | cbm71            | X94146            |
| Cry17A           | cbm71            | X99478            |
| Cry18A           | CryBP1           | X99049            |
| Cry19A           | Jeg65            | Y08920            |
| Cyt1Aa           | CytA             | X03182            |
| Cyt1Ab           | CytM             | X98793            |
| Cyt2A            | CytB             | Z14147            |
| Cyt2B            | CytB             | U52043            |

<sup>a</sup>Adapted from: [http://epunix.biols.susx.ac.uk/Home/Neil\\_Crickmore/Bt/index.html](http://epunix.biols.susx.ac.uk/Home/Neil_Crickmore/Bt/index.html)

Protease inhibitors also may provide insect resistance (Johnson *et al.*, 1989), and will thus have utility in plant transformation. The use of a protease inhibitor II gene, *pinII*, from 5 tomato or potato is envisioned to be particularly useful. Even more advantageous is the use of a *pinII* gene in combination with a Bt toxin gene, the combined effect of which has been discovered to produce synergistic insecticidal activity. Other genes which encode inhibitors of the insect's digestive system, or those that encode enzymes or co-factors that facilitate the production of inhibitors, also may be useful. This group may be exemplified by oryzacystatin 10 and amylase inhibitors such as those from wheat and barley.

Also, genes encoding lectins may confer additional or alternative insecticide properties. Lectins (originally termed phytohemagglutinins) are multivalent carbohydrate-binding proteins which have the ability to agglutinate red blood cells from a range of species. Lectins have been identified recently as insecticidal agents with activity against weevils, ECB and rootworm 15 (Murdock *et al.*, 1990; Czapla & Lang, 1990). Lectin genes contemplated to be useful include, for example, barley and wheat germ agglutinin (WGA) and rice lectins (Gatehouse *et al.*, 1984), with WGA being preferred.

Genes controlling the production of large or small polypeptides active against insects when introduced into the insect pests, such as, e.g., lytic peptides, peptide hormones and toxins and venoms, form another aspect of the invention. For example, it is contemplated that the expression of juvenile hormone esterase, directed towards specific insect pests, also may result 5 in insecticidal activity, or perhaps cause cessation of metamorphosis (Hammock *et al.*, 1990).

Transgenic plants expressing genes which encode enzymes that affect the integrity of the insect cuticle form yet another aspect of the invention. Such genes include those encoding, e.g., chitinase, proteases, lipases and also genes for the production of nikkomycin, a compound that inhibits chitin synthesis, the introduction of any of which is contemplated to produce insect 10 resistant plants. Genes that code for activities that affect insect molting, such as those affecting the production of ecdysteroid UDP-glucosyl transferase, also fall within the scope of the useful transgenes of the present invention.

Genes that code for enzymes that facilitate the production of compounds that reduce the nutritional quality of the host plant to insect pests also are encompassed by the present 15 invention. It may be possible, for instance, to confer insecticidal activity on a plant by altering its sterol composition. Sterols are obtained by insects from their diet and are used for hormone synthesis and membrane stability. Therefore alterations in plant sterol composition by expression of novel genes, e.g., those that directly promote the production of undesirable sterols or those that convert desirable sterols into undesirable forms, could have a negative effect on 20 insect growth and/or development and hence endow the plant with insecticidal activity. Lipoxygenases are naturally occurring plant enzymes that have been shown to exhibit anti-nutritional effects on insects and to reduce the nutritional quality of their diet. Therefore, further embodiments of the invention concern transgenic plants with enhanced lipoxygenase activity which may be resistant to insect feeding.

25 *Tripsacum dactyloides* is a species of grass that is resistant to certain insects, including corn root worm. It is anticipated that genes encoding proteins that are toxic to insects or are involved in the biosynthesis of compounds toxic to insects will be isolated from *Tripsacum* and that these novel genes will be useful in conferring resistance to insects. It is known that the basis of insect resistance in *Tripsacum* is genetic, because said resistance has been transferred 30 to *Zea mays* via sexual crosses (Branson and Guss, 1972). It is further anticipated that other

cereal, monocot or dicot plant species may have genes encoding proteins that are toxic to insects which would be useful for producing insect resistant plants.

Further genes encoding proteins characterized as having potential insecticidal activity also may be used as transgenes in accordance herewith. Such genes include, for example, the 5 cowpea trypsin inhibitor (CpTI; Hilder *et al.*, 1987) which may be used as a rootworm deterrent; genes encoding avermectin (*Avermectin and Abamectin*, Campbell, W.C., Ed., 1989; Ikeda *et al.*, 1987) which may prove particularly useful as a corn rootworm deterrent; ribosome inactivating protein genes; and even genes that regulate plant structures. Transgenic plants including anti-insect antibody genes and genes that code for enzymes that can convert a non-toxic insecticide (pro-insecticide) applied to the outside of the plant into an insecticide inside 10 the plant also are contemplated.

(iii) *Environment or Stress Resistance*

Improvement of a plants ability to tolerate various environmental stresses such as, but not limited to, drought, excess moisture, chilling, freezing, high temperature, salt, and oxidative 15 stress, also can be effected through expression of novel genes. It is proposed that benefits may be realized in terms of increased resistance to freezing temperatures through the introduction of an "antifreeze" protein such as that of the Winter Flounder (Cutler *et al.*, 1989) or synthetic gene derivatives thereof. Improved chilling tolerance also may be conferred through increased expression of glycerol-3-phosphate acetyltransferase in chloroplasts (Wolter *et al.*, 1992). 20 Resistance to oxidative stress (often exacerbated by conditions such as chilling temperatures in combination with high light intensities) can be conferred by expression of superoxide dismutase (Gupta *et al.*, 1993), and may be improved by glutathione reductase (Bowler *et al.*, 1992). Such strategies may allow for tolerance to freezing in newly emerged fields as well as extending later maturity higher yielding varieties to earlier relative maturity zones.

25 It is contemplated that the expression of novel genes that favorably effect plant water content, total water potential, osmotic potential, and turgor will enhance the ability of the plant to tolerate drought. As used herein, the terms "drought resistance" and "drought tolerance" are used to refer to a plants increased resistance or tolerance to stress induced by a reduction in water availability, as compared to normal circumstances, and the ability of the plant to function 30 and survive in lower-water environments. In this aspect of the invention it is proposed, for

example, that the expression of genes encoding for the biosynthesis of osmotically-active solutes, such as polyol compounds, may impart protection against drought. Within this class are genes encoding for mannitol-L-phosphate dehydrogenase (Lee and Saier, 1982) and trehalose-6-phosphate synthase (Kaasen *et al.*, 1992). Through the subsequent action of native phosphatases in the cell or by the introduction and coexpression of a specific phosphatase, these introduced genes will result in the accumulation of either mannitol or trehalose, respectively, both of which have been well documented as protective compounds able to mitigate the effects of stress. Mannitol accumulation in transgenic tobacco has been verified and preliminary results indicate that plants expressing high levels of this metabolite are able to tolerate an applied osmotic stress (Tarczynski *et al.*, 1992, 1993).

Similarly, the efficacy of other metabolites in protecting either enzyme function (*e.g.*, alanopine or propionic acid) or membrane integrity (*e.g.*, alanopine) has been documented (Loomis *et al.*, 1989), and therefore expression of genes encoding for the biosynthesis of these compounds might confer drought resistance in a manner similar to or complimentary to mannitol. Other examples of naturally occurring metabolites that are osmotically active and/or provide some direct protective effect during drought and/or desiccation include fructose, erythritol (Coxson *et al.*, 1992), sorbitol, dulcitol (Karsten *et al.*, 1992), glucosylglycerol (Reed *et al.*, 1984; Erdmann *et al.*, 1992), sucrose, stachyose (Koster and Leopold, 1988; Blackman *et al.*, 1992), raffinose (Bernal-Lugo and Leopold, 1992), proline (Rensburg *et al.*, 1993), glycine betaine, ononitol and pinitol (Vernon and Bohnert, 1992). Continued canopy growth and increased reproductive fitness during times of stress will be augmented by introduction and expression of genes such as those controlling the osmotically active compounds discussed above and other such compounds. Currently preferred genes which promote the synthesis of an osmotically active polyol compound are genes which encode the enzymes mannitol-1-phosphate dehydrogenase, trehalose-6-phosphate synthase and myoinositol 0-methyltransferase.

It is contemplated that the expression of specific proteins also may increase drought tolerance. Three classes of Late Embryogenic Proteins have been assigned based on structural similarities (see Dure *et al.*, 1989). All three classes of LEAs have been demonstrated in maturing (*i.e.* desiccating) seeds. Within these 3 types of LEA proteins, the Type-II (dehydron-type) have generally been implicated in drought and/or desiccation tolerance in vegetative plant

parts (*i.e.* Mundy and Chua, 1988; Piatkowski *et al.*, 1990; Yamaguchi-Shinozaki *et al.*, 1992). Recently, expression of a Type-III LEA (HVA-1) in tobacco was found to influence plant height, maturity and drought tolerance (Fitzpatrick, 1993). In rice, expression of the HVA-1 gene influenced tolerance to water deficit and salinity (Xu *et al.*, 1996). Expression of 5 structural genes from all three LEA groups may therefore confer drought tolerance. Other types of proteins induced during water stress include thiol proteases, aldolases and transmembrane transporters (Guerrero *et al.*, 1990), which may confer various protective and/or repair-type functions during drought stress. It also is contemplated that genes that effect lipid biosynthesis and hence membrane composition might also be useful in conferring drought resistance on the 10 plant.

Many of these genes for improving drought resistance have complementary modes of action. Thus, it is envisaged that combinations of these genes might have additive and/or synergistic effects in improving drought resistance in plants. Many of these genes also improve freezing tolerance (or resistance); the physical stresses incurred during freezing and drought are 15 similar in nature and may be mitigated in similar fashion. Benefit may be conferred via constitutive expression of these genes, but the preferred means of expressing these novel genes may be through the use of a turgor-induced promoter (such as the promoters for the turgor-induced genes described in Guerrero *et al.*, 1990 and Shagan *et al.*, 1993 which are incorporated herein by reference). Spatial and temporal expression patterns of these genes may 20 enable plants to better withstand stress.

It is proposed that expression of genes that are involved with specific morphological traits that allow for increased water extractions from drying soil would be of benefit. For example, introduction and expression of genes that alter root characteristics may enhance water uptake. It also is contemplated that expression of genes that enhance reproductive fitness 25 during times of stress would be of significant value. For example, expression of genes that improve the synchrony of pollen shed and receptiveness of the female flower parts, *i.e.*, silks, would be of benefit. In addition it is proposed that expression of genes that minimize kernel abortion during times of stress would increase the amount of grain to be harvested and hence be of value.

30 Given the overall role of water in determining yield, it is contemplated that enabling plants to utilize water more efficiently, through the introduction and expression of novel genes,

will improve overall performance even when soil water availability is not limiting. By introducing genes that improve the ability of plants to maximize water usage across a full range of stresses relating to water availability, yield stability or consistency of yield performance may be realized.

5 (iv) *Disease Resistance*

It is proposed that increased resistance to diseases may be realized through introduction of genes into plants, for example, into monocotyledonous plants such as maize. It is possible to produce resistance to diseases caused by viruses, bacteria, fungi and nematodes. It also is contemplated that control of mycotoxin producing organisms may be realized through 10 expression of introduced genes.

Resistance to viruses may be produced through expression of novel genes. For example, it has been demonstrated that expression of a viral coat protein in a transgenic plant can impart resistance to infection of the plant by that virus and perhaps other closely related viruses (Cuozzo *et al.*, 1988, Hemenway *et al.*, 1988, Abel *et al.*, 1986). It is contemplated that 15 expression of antisense genes targeted at essential viral functions may also impart resistance to viruses. For example, an antisense gene targeted at the gene responsible for replication of viral nucleic acid may inhibit replication and lead to resistance to the virus. It is believed that interference with other viral functions through the use of antisense genes also may increase resistance to viruses. Further, it is proposed that it may be possible to achieve resistance to 20 viruses through other approaches, including, but not limited to the use of satellite viruses.

It is proposed that increased resistance to diseases caused by bacteria and fungi may be realized through introduction of novel genes. It is contemplated that genes encoding so-called "peptide antibiotics," pathogenesis related (PR) proteins, toxin resistance, and proteins affecting host-pathogen interactions such as morphological characteristics will be useful. Peptide 25 antibiotics are polypeptide sequences which are inhibitory to growth of bacteria and other microorganisms. For example, the classes of peptides referred to as cecropins and magainins inhibit growth of many species of bacteria and fungi. It is proposed that expression of PR proteins in monocotyledonous plants such as maize may be useful in conferring resistance to bacterial disease. These genes are induced following pathogen attack on a host plant and have 30 been divided into at least five classes of proteins (Bol, Linthorst, and Cornelissen, 1990).

Included amongst the PR proteins are  $\beta$ -1, 3-glucanases, chitinases, and osmotin and other proteins that are believed to function in plant resistance to disease organisms. Other genes have been identified that have antifungal properties, *e.g.*, UDA (stinging nettle lectin) and hevein (Broekaert *et al.*, 1989; Barkai-Golan *et al.*, 1978). It is known that certain plant diseases are 5 caused by the production of phytotoxins. It is proposed that resistance to these diseases would be achieved through expression of a novel gene that encodes an enzyme capable of degrading or otherwise inactivating the phytotoxin. It also is contemplated that expression of novel genes that alter the interactions between the host plant and pathogen may be useful in reducing the ability of the disease organism to invade the tissues of the host plant, *e.g.*, an increase in the 10 waxiness of the leaf cuticle or other morphological characteristics.

(v) *Plant Agronomic Characteristics*

Two of the factors determining where crop plants can be grown are the average daily temperature during the growing season and the length of time between frosts. Within the areas where it is possible to grow a particular crop, there are varying limitations on the maximal time 15 it is allowed to grow to maturity and be harvested. For example, a variety to be grown in a particular area is selected for its ability to mature and dry down to harvestable moisture content within the required period of time with maximum possible yield. Therefore, crops of varying maturities are developed for different growing locations. Apart from the need to dry down sufficiently to permit harvest, it is desirable to have maximal drying take place in the field to 20 minimize the amount of energy required for additional drying post-harvest. Also, the more readily a product such as grain can dry down, the more time there is available for growth and kernel fill. It is considered that genes that influence maturity and/or dry down can be identified and introduced into plant lines using transformation techniques to create new varieties adapted to different growing locations or the same growing location, but having improved yield to 25 moisture ratio at harvest. Expression of genes that are involved in regulation of plant development may be especially useful.

It is contemplated that genes may be introduced into plants that would improve standability and other plant growth characteristics. Expression of novel genes in plants which confer stronger stalks, improved root systems, or prevent or reduce ear droppage would be of 30 great value to the farmer. It is proposed that introduction and expression of genes that increase the total amount of photoassimilate available by, for example, increasing light distribution

and/or interception would be advantageous. In addition, the expression of genes that increase the efficiency of photosynthesis and/or the leaf canopy would further increase gains in productivity. It is contemplated that expression of a phytochrome gene in crop plants may be advantageous. Expression of such a gene may reduce apical dominance, confer semidwarfism 5 on a plant, and increase shade tolerance (U.S. Patent No. 5,268,526). Such approaches would allow for increased plant populations in the field.

(vi) *Nutrient Utilization*

The ability to utilize available nutrients may be a limiting factor in growth of crop plants. It is proposed that it would be possible to alter nutrient uptake, tolerate pH extremes, 10 mobilization through the plant, storage pools, and availability for metabolic activities by the introduction of novel genes. These modifications would allow a plant such as maize to more efficiently utilize available nutrients. It is contemplated that an increase in the activity of, for example, an enzyme that is normally present in the plant and involved in nutrient utilization would increase the availability of a nutrient. An example of such an enzyme would be phytase. 15 It is further contemplated that enhanced nitrogen utilization by a plant is desirable. Expression of a glutamate dehydrogenase gene in plants, *e.g.*, *E. coli* *gdhA* genes, may lead to increased fixation of nitrogen in organic compounds. Furthermore, expression of *gdhA* in plants may lead to enhanced resistance to the herbicide glufosinate by incorporation of excess ammonia into glutamate, thereby detoxifying the ammonia. It also is contemplated that expression of a novel 20 gene may make a nutrient source available that was previously not accessible, *e.g.*, an enzyme that releases a component of nutrient value from a more complex molecule, perhaps a macromolecule.

(vii) *Male Sterility*

Male sterility is useful in the production of hybrid seed. It is proposed that male 25 sterility may be produced through expression of novel genes. For example, it has been shown that expression of genes that encode proteins that interfere with development of the male inflorescence and/or gametophyte result in male sterility. Chimeric ribonuclease genes that express in the anthers of transgenic tobacco and oilseed rape have been demonstrated to lead to male sterility (Mariani *et al.*, 1990).

A number of mutations were discovered in maize that confer cytoplasmic male sterility. One mutation in particular, referred to as T cytoplasm, also correlates with sensitivity to Southern corn leaf blight. A DNA sequence, designated TURF-13 (Levings, 1990), was identified that correlates with T cytoplasm. It is proposed that it would be possible through the 5 introduction of TURF-13 via transformation, to separate male sterility from disease sensitivity. As it is necessary to be able to restore male fertility for breeding purposes and for grain production, it is proposed that genes encoding restoration of male fertility also may be introduced.

(viii) *Negative Selectable Markers*

10 Introduction of genes encoding traits that can be selected against may be useful for eliminating undesirable linked genes. It is contemplated that when two or more genes are introduced together by cotransformation that the genes will be linked together on the host chromosome. For example, a gene encoding a *Bt* gene that confers insect resistance on the plant may be introduced into a plant together with a *bar* gene that is useful as a selectable 15 marker and confers resistance to the herbicide Liberty® on the plant. However, it may not be desirable to have an insect resistant plant that also is resistant to the herbicide Liberty®. It is proposed that one also could introduce an antisense *bar* gene that is expressed in those tissues where one does not want expression of the *bar* gene, e.g., in whole plant parts. Hence, although the *bar* gene is expressed and is useful as a selectable marker, it is not useful to confer herbicide 20 resistance on the whole plant. The *bar* antisense gene is a negative selectable marker.

It also is contemplated that negative selection is necessary in order to screen a population of transformants for rare homologous recombinants generated through gene targeting. For example, a homologous recombinant may be identified through the inactivation 25 of a gene that was previously expressed in that cell. The antisense gene to neomycin phosphotransferase II (NPT II) has been investigated as a negative selectable marker in tobacco (*Nicotiana tabacum*) and *Arabidopsis thaliana* (Xiang. and Guerra, 1993). In this example, both sense and antisense NPT II genes are introduced into a plant through transformation and the resultant plants are sensitive to the antibiotic kanamycin. An introduced gene that integrates 30 into the host cell chromosome at the site of the antisense NPT II gene, and inactivates the antisense gene, will make the plant resistant to kanamycin and other aminoglycoside antibiotics. Therefore, rare, site-specific recombinants may be identified by screening for

antibiotic resistance. Similarly, any gene, native to the plant or introduced through transformation, that when inactivated confers resistance to a compound, may be useful as a negative selectable marker.

It is contemplated that negative selectable markers also may be useful in other ways.

5 One application is to construct transgenic lines in which one could select for transposition to unlinked sites. In the process of tagging it is most common for the transposable element to move to a genetically linked site on the same chromosome. A selectable marker for recovery of rare plants in which transposition has occurred to an unlinked locus would be useful. For example, the enzyme cytosine deaminase may be useful for this purpose (Stougaard, 1993). In

10 the presence of this enzyme the compound 5-fluorocytosine is converted to 5-fluorouracil which is toxic to plant and animal cells. If a transposable element is linked to the gene for the enzyme cytosine deaminase, one may select for transposition to unlinked sites by selecting for transposition events in which the resultant plant is now resistant to 5-fluorocytosine. The parental plants and plants containing transpositions to linked sites will remain sensitive to 5-

15 fluorocytosine. Resistance to 5-fluorocytosine is due to loss of the cytosine deaminase gene through genetic segregation of the transposable element and the cytosine deaminase gene. Other genes that encode proteins that render the plant sensitive to a certain compound will also be useful in this context. For example, T-DNA gene 2 from *Agrobacterium tumefaciens* encodes a protein that catalyzes the conversion of  $\alpha$ -naphthalene acetamide (NAM) to  $\alpha$ -

20 naphthalene acetic acid (NAA) renders plant cells sensitive to high concentrations of NAM (Depicker *et al.*, 1988).

It also is contemplated that negative selectable markers may be useful in the construction of transposon tagging lines. For example, by marking an autonomous transposable element such as Ac, Master Mu, or En/Spn with a negative selectable marker, one could select

25 for transformants in which the autonomous element is not stably integrated into the genome. It is proposed that this would be desirable, for example, when transient expression of the autonomous element is desired to activate in trans the transposition of a defective transposable element, such as Ds, but stable integration of the autonomous element is not desired. The presence of the autonomous element may not be desired in order to stabilize the defective element, *i.e.*, prevent it from further transposing. However, it is proposed that if stable

30 integration of an autonomous transposable element is desired in a plant the presence of a

negative selectable marker may make it possible to eliminate the autonomous element during the breeding process.

(ix) *Non-Protein-Expressing Sequences*

5 DNA may be introduced into plants for the purpose of expressing RNA transcripts that function to affect plant phenotype yet are not translated into protein. Two examples are antisense RNA and RNA with ribozyme activity. Both may serve possible functions in reducing or eliminating expression of native or introduced plant genes. However, as detailed below, DNA need not be expressed to effect the phenotype of a plant.

1. Antisense RNA

10 Genes may be constructed or isolated, which when transcribed, produce antisense RNA that is complementary to all or part(s) of a targeted messenger RNA(s). The antisense RNA reduces production of the polypeptide product of the messenger RNA. The polypeptide product may be any protein encoded by the plant genome. The aforementioned genes will be referred to as antisense genes. An antisense gene may thus be introduced into a plant by transformation 15 methods to produce a novel transgenic plant with reduced expression of a selected protein of interest. For example, the protein may be an enzyme that catalyzes a reaction in the plant. Reduction of the enzyme activity may reduce or eliminate products of the reaction which include any enzymatically synthesized compound in the plant such as fatty acids, amino acids, carbohydrates, nucleic acids and the like. Alternatively, the protein may be a storage protein, 20 such as a zein, or a structural protein, the decreased expression of which may lead to changes in seed amino acid composition or plant morphological changes respectively. The possibilities cited above are provided only by way of example and do not represent the full range of applications.

2. Ribozymes

25 Genes also may be constructed or isolated, which when transcribed, produce RNA enzymes (ribozymes) which can act as endoribonucleases and catalyze the cleavage of RNA molecules with selected sequences. The cleavage of selected messenger RNAs can result in the reduced production of their encoded polypeptide products. These genes may be used to prepare novel transgenic plants which possess them. The transgenic plants may possess reduced levels 30 of polypeptides including, but not limited to, the polypeptides cited above.

Ribozymes are RNA-protein complexes that cleave nucleic acids in a site-specific fashion. Ribozymes have specific catalytic domains that possess endonuclease activity (Kim and Cech, 1987; Gerlach *et al.*, 1987; Forster and Symons, 1987). For example, a large number of ribozymes accelerate phosphoester transfer reactions with a high degree of specificity, often 5 cleaving only one of several phosphoesters in an oligonucleotide substrate (Cech *et al.*, 1981; Michel and Westhof, 1990; Reinhold-Hurek and Shub, 1992). This specificity has been attributed to the requirement that the substrate bind *via* specific base-pairing interactions to the internal guide sequence ("IGS") of the ribozyme prior to chemical reaction.

Ribozyme catalysis has primarily been observed as part of sequence-specific 10 cleavage/ligation reactions involving nucleic acids (Joyce, 1989; Cech *et al.*, 1981). For example, U. S. Patent 5,354,855 reports that certain ribozymes can act as endonucleases with a sequence specificity greater than that of known ribonucleases and approaching that of the DNA restriction enzymes.

Several different ribozyme motifs have been described with RNA cleavage activity 15 (Symons, 1992). Examples include sequences from the Group I self splicing introns including Tobacco Ringspot Virus (Prody *et al.*, 1986), Avocado Sunblotch Viroid (Palukaitis *et al.*, 1979; Symons, 1981), and Lucerne Transient Streak Virus (Forster and Symons, 1987). Sequences from these and related viruses are referred to as hammerhead ribozyme based on a predicted folded secondary structure.

20 Other suitable ribozymes include sequences from RNase P with RNA cleavage activity (Yuan *et al.*, 1992, Yuan and Altman, 1994, U. S. Patents 5,168,053 and 5,624,824), hairpin ribozyme structures (Berzal-Herranz *et al.*, 1992; Chowrira *et al.*, 1993) and Hepatitis Delta virus based ribozymes (U. S. Patent 5,625,047). The general design and optimization of ribozyme directed RNA cleavage activity has been discussed in detail (Haseloff and Gerlach, 25 1988, Symons, 1992, Chowrira *et al.*, 1994; Thompson *et al.*, 1995).

The other variable on ribozyme design is the selection of a cleavage site on a given 30 target RNA. Ribozymes are targeted to a given sequence by virtue of annealing to a site by complimentary base pair interactions. Two stretches of homology are required for this targeting. These stretches of homologous sequences flank the catalytic ribozyme structure defined above. Each stretch of homologous sequence can vary in length from 7 to 15

nucleotides. The only requirement for defining the homologous sequences is that, on the target RNA, they are separated by a specific sequence which is the cleavage site. For hammerhead RNA, the cleavage site is a dinucleotide sequence on the target RNA is a uracil (U) followed by either an adenine, cytosine or uracil (A,C or U) (Perriman *et al.*, 1992; Thompson *et al.*, 1995). The frequency of this dinucleotide occurring in any given RNA is statistically 3 out of 16. Therefore, for a given target messenger RNA of 1000 bases, 187 dinucleotide cleavage sites are statistically possible.

Designing and testing ribozymes for efficient cleavage of a target RNA is a process well known to those skilled in the art. Examples of scientific methods for designing and testing 10 ribozymes are described by Chowrira *et al.*, (1994) and Lieber and Strauss (1995), each incorporated by reference. The identification of operative and preferred sequences for use in down regulating a given gene is simply a matter of preparing and testing a given sequence, and is a routinely practiced "screening" method known to those of skill in the art.

### 3. Induction of Gene Silencing

15 It also is possible that genes may be introduced to produce novel transgenic plants which have reduced expression of a native gene product by the mechanism of co-suppression. It has been demonstrated in tobacco, tomato, and petunia (Goring *et al.*, 1991; Smith *et al.*, 1990; Napoli *et al.*, 1990; van der Krol *et al.*, 1990) that expression of the sense transcript of a 20 native gene will reduce or eliminate expression of the native gene in a manner similar to that observed for antisense genes. The introduced gene may encode all or part of the targeted native protein but its translation may not be required for reduction of levels of that native protein.

### 4. Non-RNA-Expressing Sequences

DNA elements including those of transposable elements such as Ds, Ac, or Mu, may be 25 inserted into a gene to cause mutations. These DNA elements may be inserted in order to inactivate (or activate) a gene and thereby "tag" a particular trait. In this instance the transposable element does not cause instability of the tagged mutation, because the utility of the element does not depend on its ability to move in the genome. Once a desired trait is tagged, the introduced DNA sequence may be used to clone the corresponding gene, *e.g.*, using the introduced DNA sequence as a PCR primer together with PCR gene cloning techniques 30 (Shapiro, 1983; Dellaporta *et al.*, 1988). Once identified, the entire gene(s) for the particular

trait, including control or regulatory regions where desired, may be isolated, cloned and manipulated as desired. The utility of DNA elements introduced into an organism for purposes of gene tagging is independent of the DNA sequence and does not depend on any biological activity of the DNA sequence, *i.e.*, transcription into RNA or translation into protein. The sole 5 function of the DNA element is to disrupt the DNA sequence of a gene.

It is contemplated that unexpressed DNA sequences, including novel synthetic sequences, could be introduced into cells as proprietary "labels" of those cells and plants and seeds thereof. It would not be necessary for a label DNA element to disrupt the function of a gene endogenous to the host organism, as the sole function of this DNA would be to identify 10 the origin of the organism. For example, one could introduce a unique DNA sequence into a plant and this DNA element would identify all cells, plants, and progeny of these cells as having arisen from that labeled source. It is proposed that inclusion of label DNAs would enable one to distinguish proprietary germplasm or germplasm derived from such, from unlabelled germplasm.

15 Another possible element which may be introduced is a matrix attachment region element (MAR), such as the chicken lysozyme A element (Stief, 1989), which can be positioned around an expressible gene of interest to effect an increase in overall expression of the gene and diminish position dependent effects upon incorporation into the plant genome (Stief *et al.*, 1989; Phi-Van *et al.*, 1990).

20 **XVIII. Examples**

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skilled the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventors to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its 25 practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or 30 similar results would be achieved. All such similar substitutes and modifications apparent to

those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

### EXAMPLE 1

#### Generation of an *Arabidopsis thaliana* Mapping Population

5 To generate a pollen donor plant, two alleles of *qr1l* were crossed to one another. The *qr1l-1* allele was in the Landsberg ecotype background and the *qr1l-2* allele was in the Columbia ecotype background. The Landsberg ecotype was readily discernible from the Columbia ecotype because it carries a recessive mutation, *erecta*, which causes the stems to thicken, inflorescences to be more compact, and the leaves to be more rounded and small than wildtype.

10 10 To utilize this as a marker of a donor plant, *qr1l-2* pollen was crossed onto a *qr1l-1* female stigma. The  $F_1$  progeny were heterozygous at all molecular markers yet the progeny retain the quartet phenotype of a tetrad of fused pollen grains. In addition, progeny display the ERECTA phenotype of the Columbia plant. This visible marker serves as an indication that the crossing was successful in generating plants segregating ecotype specific markers. Further testing was

15 15 done to the donor plants by performing PCR analysis to insure that progeny were heterozygous at molecular loci.

20 Due to the fact that the pollen grains can not be directly assayed for marker segregation and because of the desire to create a long-term resource available for multiple marker assays, it was necessary to cross individual tetrads generated by the donor plant. This created sets of progeny plants which yielded both large quantities of tissue and seed. These crosses were accomplished efficiently by generating a recipient plant homozygous for *glaborous 1-1* and male sterility. The recessive visible mutant, *glaborous1-1*, was chosen to guard against the possibility of the recipient plant self-fertilizing and the progeny being mistaken for tetrad plants. Pollen inviability through *male sterile-1* was introduced to guard against the recipient plant self-fertilizing. Due to the fact that the homozygous plant does not self, a stock seed generated by a heterozygous *male sterility 1* plant needs to be maintained from which sterile recipient plants can be selected. A large resource of seed segregating the optimal sterile and visually marked recipient plant was obtained.

**EXAMPLE 2****Tetrad Pollinations**

Tetrad pollinations were carried out as follows. A mature flower was removed from the donor plant and tapped upon a glass microscope slide to release mature tetrad pollen grains. 5 This slide was then placed under a 20-40x Zeiss dissecting microscope. To isolate individual tetrad pollen grains, a small wooden dowel was used to which an eyebrow hair with rubber cement was mounted. Using the light microscope, a tetrad pollen unit was chosen and touched to the eyebrow hair. The tetrad preferentially adhered to the eyebrow hair and was thus lifted from the microscope slide and transported the recipient plant stigmatic surface. The transfer 10 was carried out without the use of the microscope, and the eyebrow hair with adhering tetrad was then placed against the recipient stigmatic surface and the hair was manually dragged across the stigma surface. The tetrad then preferentially adhered to the stigma of the recipient and the cross pollination was completed.

Preferably, an additional backcross of the *qr1l-2* parent was used to increase the 15 pollination efficiency, which was increased to >80% successful seed set production. Initially, 57 tetrad seed sets consisting of 3-4 seeds each, were collected. Plants were grown from these tetrad seed sets, and tissue was collected. DNA was extracted from a small portion of the stored tissue for PCR based segregation analysis. Additionally the segregation of the visible *erecta* phenotype was scored. When the plants set seed, the seed was collected as a source for 20 the larger amounts of DNA required to analyze RFLP segregation by Southern blotting.

**EXAMPLE 3****Genetic Mapping of Centromeres**

To map centromeres, *F*<sub>1</sub> plants which were heterozygous for hundreds of polymorphic DNA markers were generated by crossing *quartet* mutants from the Landsberg and Columbia 25 ecotypes (Chang *et al.* 1988; Ecker, 1994; Konieczny and Ausubel, 1993). In tetrads from these plants, genetic markers segregate in a 2:2 ratio (FIG. 6; Preuss *et al.* 1994). The segregation of markers was then determined in plants which were generated by crossing pollen grains from the *F*<sub>1</sub> plants onto a Landsberg homozygote. The genotype of the pollen grains within a tetrad was inferred from the genotype of the progeny. Initially, seeds were generated from greater than 30 100 successful tetrad pollinations, and tissue and seeds were collected from 57 of these. This

provided sufficient material for PCR, as well as seeds necessary for producing the large quantities of tissue required for Southern hybridization and RFLP mapping. In order to obtain a more precise localization of the centromeres the original tetrad population was increased from 57 tetrads to over 388 tetrads. Additional tetrads may be collected to provide even better 5 resolution.

PCR analysis was performed to determine marker segregation. To account for the contribution of the Landsberg background from the female parent, one Landsberg complement from each of the four tetrad plants was subtracted. As shown in FIG. 5, markers from sites spanning the entire genome were used for pair-wise comparisons of all other markers. 10 Tetratypes indicate a crossover between one or both markers and their centromeres whereas ditypes indicate the absence of crossovers (or presence of a double crossover).

Thus, at every genetic locus, the resulting diploid progeny was either L/C or C/C. The map generated with these plants is based solely on male meioses, unlike the existing map, which represents an average of recombination's in both males and females. Therefore, several 15 well-established genetic distances were recalculated and thus will determine whether recombination frequencies are significantly altered.

The large quantities of genetic data generated by the analysis must be compared pair-wise to perform tetrad analysis. All of the data was managed in a Microsoft Excel spread-sheet format, assigning Landsberg alleles a value of "1" and Columbia alleles a value of "0". Within 20 a tetrad, the segregation of markers on one chromosome was compared to centromere-linked reference loci on a different chromosome (see Table 2 below). Multiplying the values of each locus by an appropriate reference, and adding the results for each tetrad easily distinguished PD, NPD, and TT tetrads with values of 2, 0, and 1, respectively. By scoring more than 53 PCR-based genetic markers distributed across the genome, all five *Arabidopsis thaliana* 25 centromeres were mapped to small intervals (FIG. 3A-3E). Additionally, for each centromeric interval, a number of useful recombinants were identified.

Table 2  
Scoring protocol for tetratypes.

| Individual |       | members<br>of a tetrad | Reference |         | Reference |         | Reference |  |
|------------|-------|------------------------|-----------|---------|-----------|---------|-----------|--|
| Locus 1    | Locus |                        | Locus 2   | Locus   | Locus 3   | Locus   | Locus     |  |
| A          | 1     | x 1 = 1                | 0         | x 1 = 0 | 0         | x 1 = 0 |           |  |
| B          | 1     | x 1 = 1                | 0         | x 1 = 0 | 1         | x 1 = 1 |           |  |
| C          | 0     | x 0 = 0                | 1         | x 0 = 0 | 0         | x 0 = 0 |           |  |
| D          | 0     | x 0 = 0                | 1         | x 0 = 0 | 1         | x 0 = 0 |           |  |
| -          |       | 2                      |           | 0       |           | 1       |           |  |
|            |       | PD                     |           | NPD     |           | TT      |           |  |

#### EXAMPLE 4

##### Mapping Results: *Arabidopsis* Chromosomes 1-5

5 The centromere on chromosome 1 was mapped between mi342 (56.7 cM) and T27K12 (59.1 cM). A more refined position places the centromere between the marker T22C23 (~58.5 cM) and T27K12 (59.1 cM). However, since T22C23 is not precisely mapped the centromere may be between the publicly available markers mi342 and T27K12.

10 The centromere on chromosome 2 was mapped between mi310 (18.6 cM) and g4133 (23.8 cM). Within this interval the centromere may be between the following publicly available marker pairs: mi310 and mi421, mi421 and g4532, g4532 and SEP2A, and SEP2A and g4133.

15 The centromere on chromosome 3 was mapped between atpox (48.6 cM) and ve021 (54.7 cM). A more refined position places the centromere between the marker atpox (48.6 cM) and 91F1T7 (~54.2 cM) however, since 91F1T7 is not precisely mapped, the centromere may be between the following publicly available marker pairs: atpox and zim2, zim2 and mi79b, mi79b and RCEN3, RCEN3 and ASD, ASD and a-1, a-1 and t04109, and t04109 and ve021.

20 The centromere on chromosome 4 was mapped between mi233 (18.8 cM) and mi167 (21.5 cM). A more refined position places the centromere between the markers mi233 and F13H14 however since F13H14 is not precisely mapped, the centromere may be between the

following publicly available marker pairs: mi233 and g2616, g2616 and mi306, mi306 and m506, m506 and nga12, nga12 and BIO200, BIO200 and mi87, mi87 and C11Ath, C11Ath and m456A, and m456A and mi167.

5 The centromere on chromosome 5 was mapped between nga76 (71.6 cM) and PhC (74.3 cM). Within this interval the centromere may be between the following publicly available marker pairs: nga76 and mi291b, mi291b and CMs1, and CMs1 and PhyC.

10 All of the above markers and genetic positions (*i.e.* cM values) correspond to the Lister and Dean Recombinant Inbred Genetic map, available on-line at: [http://genome-www3.stanford.edu/atdb\\_welcome.html](http://genome-www3.stanford.edu/atdb_welcome.html).

#### EXAMPLE 5

##### Construction of Artificial Plant Chromosomes

15 Plant artificial chromosomes are constructed by combining the previously isolated essential chromosomal elements. Exemplary artificial chromosomes include those designed to be "shuttle vectors"; *i.e.*, they can be maintained in a convenient host (such as *E. coli*, *Agrobacterium* or yeast) as well as plant cells.

20 An artificial chromosome can be maintained in *E. coli* or other bacterial cells as a circular molecule by placing a removable stuffer fragment between the telomeric sequence blocks. The stuffer fragment is a dispensable DNA sequence, bordered by unique restriction sites, which can be removed by restriction digestion of the circular DNAs to create linear molecules with telomeric ends. The linear PLAC can then be isolated by, for example, gel electrophoresis. In addition to the stuffer fragment and the plant telomeres, the artificial chromosome contains a replication origin and selectable marker that can function in plants to allow the circular molecules to be maintained in bacterial cells. The artificial chromosomes also include a plant selectable marker, a plant centromere, and a plant ARS to allow replication and maintenance of the DNA molecules in plant cells. Finally, the artificial chromosome includes several unique restriction sites where additional DNA sequence inserts can be cloned. 25 The most expeditious method of physically constructing such an artificial chromosome, *i.e.*, ligating the various essential elements together for example, will be apparent to those of ordinary skill in this art.

A number of artificial chromosome vectors have been designed by the current inventors and are disclosed herein for the purpose of illustration (FIGs. 7A-7H). These vectors are not limiting however, as it will be apparent to those of skill in the art that many changes and alterations may be made and still obtain a functional vector.

5 All of the compositions and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and methods and in the steps or in the sequence of steps of the method described  
10 herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended  
15 claims.

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

- 5 Abdullah *et al.*, *Biotechnology*, 4:1087, 1986.
- Abel *et al.*, *Science*, 232:738-743, 1986.
- Alfenito *et al.*, "Molecular characterization of a maize B chromosome centric sequence," *Genetics*, 135:589-597, 1993.
- Barkai-Golan *et al.*, *Arch. Microbiol.*, 116:119-124, 1978.
- 10 Baum *et al.*, "The centromeric K-type repeat and the central core are together sufficient to establish a functional *Schizosaccharomyces pombe* centromere," *Mol. Bio. Cell.*, 5:747-761, 1994.
- Bell *et al.*, "Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*," *Genomics*, 19:137-144, 1994.
- 15 Bernal-Lugo and Leopold, *Plant Physiol.*, 98:1207-1210, 1992.
- Berzal-Herranz *et al.*, *Genes and Devel.*, 6:129-134, 1992.
- Blackman *et al.*, *Plant Physiol.*, 100:225-230, 1992.
- Bloom, "The centromere frontier: Kinetochore components, microtubule-based motility, and the CEN-value paradox," *Cell*, 73:621-624, 1993.
- 20 Bol *et al.*, *Annu. Rev. Phytopath.*, 28:113-138, 1990.
- Bowler *et al.*, *Ann Rev. Plant Physiol.*, 43:83-116, 1992.
- Branson and Guss, *Proceedings North Central Branch Entomological Society of America*, 27:91-95, 1972.
- Brisson *et al.*, *Nature*, 310:511, 1984.
- 25 Broach *et al.*, *Gene*, 8:121-133, 1979.
- Broakaert *et al.*, *Science*, 245:1100-1102, 1989.
- Burke *et al.*, *Science*, 236:806-812, 1987.
- Bytebier *et al.*, *Proc. Natl Acad. Sci. USA*, 84:5345, 1987.
- Callis *et al.*, *Genes and Development*, 1:1183, 1987.
- 30 Campbell (ed.), "In: *Avermectin and Abamectin*," 1989.

Campbell, "Monoclonal Antibody Technology, Laboratory Techniques in Biochemistry and Molecular Biology," Vol. 13, Burden and Von Knippenberg, Eds. pp. 75-83, Elsevier, Amsterdam, 1984.

5 Capecchi, "High efficiency transformation by direct microinjection of DNA into cultured mammalian cells," *Cell* 22(2):479-488, 1980.

Carbon *et al.*, *In: Recombinant Molecules Impact on Science and Society*, (Raven Press), 335-378, 1977.

Carbon *et al.*, "Centromere structure and function in budding and fission yeasts," *New Biologist*, 2:10-19, 1990.

10 Carpenter *et al.*, "The control of the distribution of meiotic exchange in *Drosophila melanogaster*," *Genetics*, 101:81-90, 1982.

Cech *et al.*, "In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence," *Cell*, 27:487-496, 1981.

15 Chang *et al.*, "Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*," *Proc. Natl. Acad. Sci. USA*, 85:6856-6860, 1988.

Chepko, *Cell*, 37:1053, 1984.

Chowrira *et al.*, "In vitro and in vivo comparison of hammerhead, hairpin, and hepatitis delta virus self-processing ribozyme cassettes," *J. Biol. Chem.*, 269:25856-25864, 1994.

20 Chu *et al.*, "Separation of large DNA molecules by contour-clamped homogeneous electric fields," *Science*, 234, 1582-1585, 1986.

Clapp, "Somatic gene therapy into hematopoietic cells. Current status and future implications," *Clin. Perinatol.*, 20(1):155-168, 1993.

Clarke *et al.*, "Isolation of a yeast centromere and construction of functional small circular 25 chromosomes," *Nature*, 287:504-509, 1980.

Cohen *et al.*, *Proc. Natl. Acad. Sci. USA*, 70:3240, 1973.

Copenhaver *et al.*, "Use of RFLPs larger than 100 kbp to map position and internal organization of the nucleolus organizer region on chromosome 2 in *Arabidopsis thaliana*," *Plant J.*, 7, 273-286, 1995.

30 Coxson *et al.*, *Biotropica*, 24:121-133, 1992.

Cristou *et al.*, *Plant Physiol.*, 87:671-674, 1988.

Cuozzo *et al.*, *Bio/Technology*, 6:549-553, 1988.

Curiel *et al.*, "Adenovirus enhancement of transferrin-polylysine-mediated gene delivery," *Proc. Natl. Acad. Sci. USA*, 88(19):8850-8854, 1991.

Curiel *et al.*, high-efficiency gene transfer mediated by adenovirus coupled to DNA-polylysine complexes," *Hum. Gen. Ther.*, 3(2):147-154, 1992.

Cutler *et al.*, *J. Plant Physiol.*, 135:351-354, 1989.

Czapla and Lang, *J. Econ. Entomol.*, 83:2480-2485, 1990.

Davies *et al.*, *Plant Physiol.*, 93:588-595, 1990.

Dellaporta *et al.*, "In: *Chromosome Structure and Function: Impact of New Concepts*, 18th Stadler Genetics Symposium," 11:263-282, 1988.

Depicker *et al.*, *Plant Cell Reports*, 7:63-66, 1988.

Dillon *et al.*, *Recombinant DNA Methodology*, 1985.

Dure *et al.*, *Plant Molecular Biology*, 12:475-486, 1989.

Earnshaw *et al.*, "Proteins of the inner and outer centromere of mitotic chromosomes," *Genome*, 31:541-552, 1989.

Earnshaw, "When is a centromere not a kinetochore?", *J. Cell Sci.*, 99:1-4, 1991.

Ecker, JR, *Genomics*, 19:137-144

Eglitis *et al.*, "Retroviral vectors for introduction of genes into mammalian cells," *Biotechniques*, 6(7):608-614, 1988.

Eglitis *et al.*, "Retroviral-mediated gene transfer into hemopoietic cells," *Adv. Exp. Med. Biol.*, 241:19-27, 1988.

Erdmann *et al.*, *J. Gen. Microbiology*, 138:363-368, 1992.

Fitzpatrick, *Gen. Engineering News*, 22:7, 1993.

Forster and Symons, "Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites," *Cell*, 49:211-220, 1987.

Fraley *et al.*, *Biotechnology*, 3:629, 1985.

Fromm *et al.*, "Expression of genes transferred into monocot and dicot plant cells by electroporation," *Proc. Natl. Acad. Sci. USA*, 82(17):5824-5828, 1985.

Fromm M.E., Taylor L.P., Walbot V. (1986). *Nature* 312:791-793.

Fujimura *et al.*, *Plant Tissue Culture Letters*, 2:74, 1985.

Fynan *et al.*, "DNA vaccines: protective immunizations by parenteral, mucosal, and gene gun inoculations," *Proc. Natl. Acad. Sci. USA*, 90(24):11478-11482, 1993.

Gatehouse *et al.*, *J. Sci. Food. Agric.*, 35:373-380, 1984.

Gefter *et al.*, *Somatic Cell Genet.*, 3:231-236, 1977.

5 Gerlach *et al.*, "Construction of a plant disease resistance gene from the satellite RNA of tobacco rindspot virus," *Nature (London)*, 328:802-805, 1987.

Goding, "Monoclonal Antibodies: Principles and Practice," pp. 60-74. 2nd Edition, Academic Press, Orlando, FL, 1986.

Goring *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:1770-1774, 1991.

10 Graham *et al.*, "Transformation of rat cells by DNA of human adenovirus 5," *Virology*, 54(2):536-539, 1973.

Guerrero *et al.*, *Plant Molecular Biology*, 15:11-26, 1990.

Gupta *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:1629-1633, 1993.

Haaf *et al.*, "Integration of human  $\alpha$ -satellite DNA into simian chromosomes: centromere 15 protein binding and disruption of normal chromosome segregation," *Cell*, 70:681-696, 1992.

Hadlaczky *et al.*, "Centromere formation in mouse cells cotransformed with human DNA and a dominant marker gene," *Proc. Natl. Acad. Sci. USA*, 88:8106-8110, 1991.

Hammock *et al.*, *Nature*, 344:458-461, 1990.

20 Hegemann *et al.*, "The centromere of budding yeast," *Bioassays*, 15(7):451-460, 1993.

Hemenway *et al.*, *The EMBO J.*, 7:1273-1280, 1988.

Hilder *et al.*, *Nature*, 330:160-163, 1987.

Hsiao *et al.*, *J. Proc. Natl. Acad. Sci. USA*, 76:3829-3833, 1979.

25 Hwang *et al.*, "Identification and map position of YAC clones comprising one-third of the *Arabidopsis* genome," *The Plant Journal*, 1:367-374, 1991.

Ikeda *et al.*, *J. Bacteriol.*, 169:5615-5621, 1987.

Johnston *et al.*, "Gene gun transfection of animal cells and genetic immunization," *Methods Cell. Biol.*, 43(A):353-365, 1994.

Jones, *Embo J.*, 4:2411-2418, 1985.

30 Jones, *Mol. Gen. Genet.*, 207:478, 1987.

Jorgensen *et al.*, *Mol. Gen. Genet.*, 207:471, 1987.

Joyce, "RNA evolution and the origins of life," *Nature*, 338:217-244, 1989.

Kaasen *et al.*, *J. Bacteriology*, 174:889-898, 1992.

Karsten *et al.*, *Botanica Marina*, 35:11-19, 1992.

Kim and Cech, "Three dimensional model of the active site of the self-splicing rRNA precursor of Tetrahymena," *Proc. Natl. Acad. Sci. USA*, 84:8788-8792, 1987.

5 Klee *et al.*, *Bio/Technology* 3:637-642, 1985.

Klein *et al.*, *Nature*, 327:70-73, 1987.

Klein *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:8502-8505, 1988.

Kohler *et al.*, *Eur. J. Immunol.*, 6:511-519, 1976.

10 Kohler *et al.*, *Nature*, 256:495-497, 1975.

Konieczny *et al.*, "A procedure for mapping *Arabidopsis* mutations using codominant ecotype-specific PCR-based markers," *The Plant Journal*, 4:403-410, 1993.

Koornneef *et al.*, *Genetica*, 61:41-46, 1983.

Koornneef, "Linkage map of *Arabidopsis thaliana* (2n=10)," *In SJ O'Brien, ed, Genetic Maps 1987: A compilation of linkage and restriction maps of genetically studied organisms*, 724-745, 1987.

15 Koornneef, "The use of telotrisomics for centromere mapping in *Arabidopsis thaliana* (L.) Heynh," *Genetica*, 62:33-40, 1983.

Koster and Leopold, *Plant Physiol.*, 88:829-832, 1988.

20 Kuby, J., *Immunology 2nd Edition*, W. H. Freeman & Company, NY, 1994

Kyte *et al.*, "A simple method for displaying the hydropathic character of a protein," *J. Mol. Biol.*, 157(1):105-132, 1982.

Lechner *et al.*, "A 240 kd multisubunit protein complex, CBF3 is a major component of the budding yeast centromere," *Cell*, 64:717-725, 1991.

25 Lee and Saier, *J. of Bacteriol.*, 153:685, 1983.

Levings, *Science*, 250:942-947, 1990.

Lewin, *Genes II*, John Wiley & Sons, Publishers, N.Y., 1985.

Li *et al.*, *P.N.A.S.*, 87:4580-4584, 1990.

Lieber and Strauss, "Selection of efficient cleavage sites in target RNAs by using a ribozyme expression library," *Mol. Cell. Biol.*, 15: 540-551, 1995.

30 Loomis *et al.*, *J. Expt. Zoology*, 252:9-15, 1989.

Lorz *et al.*, *Mol. Gen. Genet.*, 199:178, 1985.

Lu *et al.*, "High efficiency retroviral mediated gene transduction into single isolated immature and replatable CD34(3+) hematopoietic stem/progenitor cells from human umbilical cord blood," *J. Exp. Med.*, 178(6):2089-2096, 1993.

5 Maloy, S. R., "Experimental Techniques in Bacterial Genetics" Jones and Bartlett Prokop, A., and Bajpai, R. K. "Recombinant DNA Technology I" *Ann. N. Y. Acad. Sci.*, vol. 646, 1991.

Maluszynaska *et al.*, "Molecular cytogenetics of the genus *Arabidopsis*: *In situ* localization of rDNA sites, chromosome numbers and diversity in centromeric heterochromatin,"

10 *Annals Botany*, 71:479-484, 1993.

Maluszynska *et al.*, "Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana*," *Plant Jour.*, 1(2):159-166, 1991.

Maniatis *et al.*, "Molecular Cloning: a Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, NY., 1982.

15 Marcotte *et al.*, *Nature*, 335:454, 1988.

Mariani *et al.*, *Nature*, 347:737-741, 1990.

Martinez-Zapater *et al.*, *Mol. Gen. Genet.*, 204:417-423, 1986.

McCabe *et al.*, *Biotechnology*, 6:923, 1988.

Michel and Westhof, "Modeling of the three-dimensional architecture of group I catalytic

20 introns based on comparative sequence analysis," *J. Mol. Biol.*, 216:585-610, 1990.

Mortimer *et al.*, "Genetic mapping in *Saccharomyces cerevisiae*," *Life Cycle and Inheritance, In: The Molecular Biology of the Yeast Saccharomyces*, 11-26, 1981.

Mundy and Chua, *The EMBO J.*, 7:2279-2286, 1988.

Murdock *et al.*, *Phytochemistry*, 29:85-89, 1990.

25 Murray *et al.*, *Nature*, 305:189-193, 1983.

Napoli, Lemieux, Jorgensen, "Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes *in trans*," *Plant Cell*, 2:279-289, 1990.

Nester, *Ann. Rev. Plant Phys.*, 35:387-413, 1984.

30 Nicklas, "The forces that move chromosomes in mitosis," *Annu. Rev. Biophys. Biophys. Chem.*, 17:431-39, 1988.

Nussbaum *et al.*, *Proc. Natl. Acad. Sci USA*, 73:1068, 1976.

Omirulleh *et al.*, *Plant Molecular Biology*, 21:415-428, 1993.

Palukaitis *et al.*, "Characterization of a viroid associated with avocado sunblotch disease," *Virology*, 99:145-151, 1979.

5 Perkins, "The detection of linkage in tetrad analysis," *Genetics*, 38, 187-197, 1953.

Perlak *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:3324-3328, 1991.

Perriman *et al.*, "Extended target-site specificity for a hammerhead ribozyme," *Gene*, 113:157-163, 1992.

Phi-Van *et al.*, *Mol. Cell. Biol.*, 10:2302-2307. 1990.

10 Piatkowski *et al.*, *Plant Physiol.*, 94:1682-1688, 1990.

Potrykus *et al.*, *Mol. Gen. Genet.*, 199:183, 1985.

Preuss *et al.*, "Tetrad analysis possible in *Arabidopsis* with mutation of the QUARTET (QRT) genes," *Science*, 264:1458, 1994.

15 Price *et al.*, "In: Systematic relationships of *Arabidopsis*: a molecular and morphological perspective", Somerville, C. and Meyerowitz, E. (eds.), *Arabidopsis*, Cold Spring Harbor Press, NY, 1995.

Prody *et al.*, "Autolytic processing of dimeric plant virus satellite RNA." *Science*, 231:1577-1580, 1986.

20 Prokop *et al.*, *Ann. N. Y. Acad. Sci.*, 646, 1991

Rattner, "The structure of the mammalian centromere," *Bioassays*, 13(2):51-56, 1991.

Reed *et al.*, *J. Gen. Microbiology*, 130:1-4, 1984.

25 Reinhold-Hurek and Shub, "Self-splicing introns in tRNA genes of widely divergent bacteria," *Nature*, 357:173-176, 1992.

Rensburg *et al.*, *J. Plant Physiol.*, 141:188-194, 1993.

Richards *et al.*, "The centromere region of *Arabidopsis thaliana* chromosome 1 contains telomere-similar sequences," *Nucleic Acids Research*, 19(12):3351-3357, 1991.

30 Rieder, "The formation, structure and composition of the mammalian kinetochore and kinetochore fiber," *Int. Rev. Cytol.*, 79:1-58, 1982.

Rogers *et al.*, *Meth. in Enzymol.*, 153:253-277, 1987.

Rosenberg *et al.*, "RFLP subtraction: A method for making libraries of polymorphic markers," *Proc. Natl. Acad. Sci. USA*, 91:6113-6117, 1994.

Schwartz *et al.*, Cold Spring Harbor Symp. *Quant. Biol.*, 47, 195-198, 1982.

Sears *et al.*, "Cytogenetic studies in *Arabidopsis thaliana*," *Can. J. Genet. Cytol.*, 12:217-233, 1970.

Segal, "Biochemical Calculations," 2nd Edition. John Wiley & Sons, New York, 1976.

5 Setlow *et al.*, *Genetic Engineering: Principles and Methods*, 1979.

Shagan and Bar-Zvi, *Plant Physiol.*, 101:1397-1398, 1993.

Shapiro, "In: *Mobile Genetic Elements*," 1983.

Shingo *et al.*, *Mol. Cell. Biol.*, 6:1787, 1986.

Smith, Watson, Bird, Ray, Schuch, Grierson, "Expression of a truncated tomato 10 polygalacturonase gene inhibits expression of the endogenous gene in transgenic plants," *Mol. Gen. Genet.*, 224:447-481, 1990.

Smithies *et al.*, *Nature*, 317:230-234, 1985.

Smythe, "Pollen clusters," *Current Biology*, 4:851-853, 1994.

Spielmann *et al.*, *Mol. Gen. Genet.*, 205:34, 1986.

15 Stiefel *et al.*, *Nature*, 341:343, 1989.

Stinchcomb *et al.*, *Nature*, 282:39-43, 1979.

Stougaard, *The Plant Journal*, 3:755-761, 1993.

Symington *et al.*, *Cell*, 52:237-240, 1988.

Symons, "Avacado sunblotch viroid: primary sequence and proposed secondary structure." 20 *Nucl. Acids Res.*, 9:6527-6537, 1981.

Symons, "Small catalytic RNAs." *Annu. Rev. Biochem.*, 61:641-671, 1992.

Tarczynski *et al.*, "Expression of a bacterial *mtlD* gene in transgenic tobacco leads to 25 production and accumulation of mannitol," *Proc. Natl. Acad. Sci. USA*, 89:1-5, 1992.

Tarczynski *et al.*, "Stress Protection of Transgenic Tobacco by Production of the Osmolyte Mannitol," *Science*, 259:508-510, 1993.

Thomas *et al.*, *Cell*, 44:419-428, 1986.

Thomas *et al.*, *Proc. Natl. Acad. Sci. USA*, 71:4579, 1974.

Thompson *et al.*, "Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression," *Nature Genet.*, 9:444-450, 1995.

30 Toriyama *et al.*, *Theor Appl. Genet.*, 73:16, 1986.

Tyler-Smith *et al.*, "Mammalian chromosome structure," *Current Biology*, 3:390-397, 1993.

Uchimiya *et al.*, *Mol. Gen. Genet.*, 204:204, 1986.

Van der Krol, Mur, Beld, Mol, Stuitje, "Flavonoid genes in petunia: addition of a limiting number of copies may lead to a suppression of gene expression," *Plant Cell*, 2:291-99, 1990.

5 Vasil *et al.*, "Herbicide-resistant fertile transgenic wheat plants obtained by microprojectile bombardment of regenerable embryogenic callus," *Biotechnology*, 10:667-674, 1992.

Vasil, *Biotechnology*, 6:397, 1988.

Vernon and Bohnert, *The EMBO J.*, 11:2077-2085, 1992.

10 Wagner *et al.*, "Coupling of adenovirus to transferrin-polylysine/DNA complexes greatly enhances receptor-mediated gene delivery and expression of transfected genes," *Proc. Natl. Acad. Sci. USA* 89 (13):6099-6103, 1992.

Watrud *et al.*, In: *Engineered Organisms and the Environment*, 1985.

Watson *et al.*, *Recombinant DNA: A Short Course*, 1983.

Weinsink *et al.*, *Cell*, 3:315-325, 1974.

15 Wevrick *et al.*, "Partial deletion of alpha satellite DNA association with reduced amounts of the centromere protein CENP-B in a mitotically stable human chromosome rearrangement," *Mol Cell Biol.*, 10:6374-6380, 1990.

Whitehouse, *Nature*, No. 4205: 893, 1950.

Wigler *et al.*, *Cell*, 11:223, 1977.

20 Willard, H., "Centromeres of mammalian chromosomes," *Trends Genet.*, 6:410-416, 1990.

Willard, H., *Nature Genetics* 15:345-354, 1997

Wolter *et al.*, *The EMBO J.*, 4685-4692, 1992.

Wong *et al.*, "Electric field mediated gene transfer," *Biochim. Biophys. Res. Commun.*, 107(2):584-587, 1982.

25 Xiang and Guerra, *Plant Physiol.*, 102:287-293, 1993.

Xu *et al.*, *Plant Physiol.*, 110:249-257, 1996.

Yamada *et al.*, *Plant Cell Rep.*, 4:85, 1986.

Yamaguchi-Shinozaki *et al.*, *Plant Cell Physiol.*, 33:217-224, 1992.

Yen, *Embo J.*, 10(5), 1245-1254, 1991.

30 Young *et al.*, "In: *Eukaryotic Genetic Systems ICN-UCLA Symposia on Molecular and Cellular Biology, VII*," 315-331, 1977.

Yuan and Altman, "Selection of guide sequences that direct efficient cleavage of mRNA by human ribonuclease P," *Science*, 263:1269-1273, 1994.

Yuan *et al.*, "Targeted cleavage of mRNA by human RNase P," *Proc. Natl. Acad. Sci. USA*, 89:8006-8010, 1992.

5 Zatloukal *et al.*, "Transferrinfection: a highly efficient way to express gene constructs in eukaryotic cells," *Ann. N.Y. Acad. Sci.*, 660:136-153, 1992.

SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

(A) NAME: Arch Development Corporation  
(B) STREET: 1101 East 58th Street  
(C) CITY: Chicago  
(D) STATE: IL  
(E) COUNTRY: USA  
(F) POSTAL CODE (ZIP): 60637

(A) NAME: Greg Copenhaver  
(B) STREET: 1028 S. Elmwood  
(C) CITY: Oakpark  
(D) STATE: IL  
(E) COUNTRY: USA  
(F) POSTAL CODE (ZIP): 60304

(A) NAME: Daphne Preuss  
(B) STREET: 5807 S. Dorchester #2E  
(C) CITY: Chicago  
(D) STATE: IL  
(E) COUNTRY: USA  
(F) POSTAL CODE (ZIP): 60637

## (ii) TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS

## (iii) NUMBER OF SEQUENCES: 44

## (iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

## (vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/073,741  
(B) FILING DATE: 05-FEB-1998

## (vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/048,451  
(B) FILING DATE: 03-JUN-1997

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

76

CGCCAAAGAC TACGAAATGA TC

22

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

ATAATAGATA AAGAGCCCCA CAC

23

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGGTCTGGTT ATGCCGTGAA G

21

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GTTTTACTTA GTCCAATGGT AG

22

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

AAATGGCCAA CGATCAGAAG ATA

24

## (2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GAAGTCCGGC ATGTTATCAC CCAAG

25

## (2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CAAGTCGCAA ACGGAAAATG

20

## (2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

AAACTACGCC TAACCACTAT TCTC

24

## (2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GAAGTACAGC GGCTAAAAAA GAAG

24

## (2) INFORMATION FOR SEQ ID NO: 10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

TTGCTGCCAT GTAATACCTA AGTG

24

## (2) INFORMATION FOR SEQ ID NO: 11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GGCTACTGGT CAAATCAAAT CATTC

25

## (2) INFORMATION FOR SEQ ID NO: 12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GAATCTTGC AACCGAGTGG

20

## (2) INFORMATION FOR SEQ ID NO: 13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GCGGCTGGAT GATCTCCACC TC

22

## (2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TTACCCCGCA GGAAAAAGTA TG

22

## (2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

ACTTCATCAC TTGCAGGGACT G

21

## (2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GGCCCAAGAA GCCCACAACA C

21

## (2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GAGGATGTCC AAGGCGTCTG TGC

23

80

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GGGGATTCG GTTGTCTGTC

20

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TAACGTTCCC ACATGAGC

18

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

AACTCTGTAC CTGCTGGA

18

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

AGTCGATGTC TAGGCTCTTC

20

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

CTTCCATTTC TTGATTTAGT TC

22

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

ACTAAGGCCT GTGTTGATGT TTCTC

25

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

AACCGCTTCC CATTCTGTCTT C

21

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GGCGACCTTG GACCTGTATA CG

22

82

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

AACCGCCATT TTCATTTCTA TC

22

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

TCTGCGGTGG GAATACAAAAA G

21

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GCCATCATTC CCCGGTTCTC ATAAG

25

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

CCCCCTCCCGC CCTAAACCTA C

21

83

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

TTCCGCTACA TGGCCTTCTA CCTTG

25

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CGTATTCCCC TGAAAAGTGA CCTG

24

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

ACATCCGGCC TTCCCATTG

19

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

TTTGCACCGC CTATGTTACC

20

## (2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GAGGACGTTT TGCAGAGTG

19

## (2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

AGTGGACGCC TTCTTCAATG TG

22

## (2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

TGGTCCGTCG TAGGGCAAC

19

## (2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

CAAAACCAAA TCCGCGAAGA AC

22

85

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

AGTGGCCAGC CTTCTTAACA TACC

24

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CCGGGGAAGA AAGCGTGAAT C

21

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CCCGGAGTTA CAGCCCTTAT GATG

24

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

TTGGGGGATT GGTCAAGAAG

19

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

ATATGTTGCA ACTTAGAATC AG

22

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

TCTCGTTCTG ATGGCTCCTG TG

22

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GTGTAACCGG TGATACTCTC TCGCC

25

CLAIMS

1. A recombinant DNA construct comprising a functional *Arabidopsis thaliana* centromere.
2. The recombinant DNA construct of claim 1, which additionally comprises a telomere functional in plants.  
5
3. The recombinant DNA construct of claim 2, wherein the telomere is an *Arabidopsis thaliana* telomere.
4. The recombinant DNA construct of claim 2, wherein the telomere is a yeast telomere.
5. The recombinant DNA construct of claim 1, which additionally comprises an autonomous replicating sequence (ARS) functional in plants.  
10
6. The recombinant DNA construct of claim 5, wherein said ARS is an *Arabidopsis thaliana* ARS.
7. The recombinant DNA construct of claim 1, which additionally comprises a selectable marker gene.  
15
8. The recombinant DNA construct of claim 7, which is capable of being maintained as a chromosome, wherein said chromosome is stably transmitted in dividing cells.
9. A plasmid comprising the recombinant DNA construct of claim 7.
10. The plasmid of claim 9, wherein said plasmid further comprises an origin of replication and a selection marker that function in bacteria.
- 20 11. The plasmid of claim 10, wherein said bacteria is *E. coli*.
12. The plasmid of claim 9, wherein said plasmid further comprises an origin of replication and a selection marker that function in yeast.

13. The plasmid of claim 12, wherein said yeast is *S. cerevisiae*.
14. The recombinant DNA construct of claim 7, which additionally comprises a first structural gene.
15. The recombinant DNA construct of claim 14, further defined as comprising a second structural gene.
16. The recombinant DNA construct of claim 15, further defined as comprising a third structural gene.
17. The recombinant DNA construct of claim 16, further defined as comprising a fourth structural gene.
- 10 18. The construct of claim 14, wherein said structural gene is selected from the group consisting of an antibiotic resistance gene; a herbicide resistance gene, a nitrogen fixation gene; a plant pathogen defense gene; a plant stress-induced gene; a toxin gene; and a seed storage gene.
19. The construct of claim 18, wherein said construct is capable of expressing said desired gene sequence.
20. The construct of claim 19, wherein said construct is capable of expressing said desired gene in a prokaryote.
21. The construct of claim 19, wherein said construct is capable of expressing said desired gene in a eukaryote.
- 20 22. The construct of claim 21, wherein said eukaryote is a higher eukaryote.
23. The construct of claim 22, wherein said higher eukaryote is a plant.